

Where next with nuclear weapons?

The Non-Proliferation Treaty may collapse at the review conference in 1995 unless several urgent problems are dealt with before then. Either way, attention should shift from the avoidance of weapons to the avoidance of war.

EIGHTEEN months is but an instant by the timescales of diplomacy, but that is how long the world's nuclear powers have left to reach a deal among themselves on the disposition and development of their nuclear arsenals. Or, more precisely, if the nuclear powers are not by then at one, there is a high chance that the Nuclear Non-Proliferation Treaty (NPT) will collapse and that some of its less enthusiastic signatories will join those who never signed the treaty in seeking nuclear weapons for themselves. The circumstances, after all, are clear. The NPT, which came into effect in 1970, requires that its perpetuation beyond 25 years should then be affirmed by its members. The outlook at the last routine review conference in 1990 was bad enough; the participants could not even agree on a statement of their mutual disagreements. Since then, the basis for restraint on nuclear proliferation has been made even more precarious.

Several tendencies contribute to this gloomy prognosis. It used to be that, when the members of the treaty met every five years, the non-nuclear powers would complain that the nuclear powers had made only token efforts to implement their promises of strategic arms control. Despite the arms control measures of the past several years and, more important, the apparent ending of the Cold War whose central feature was the threat that the two most powerful nuclear powers would annihilate each other if they chose, complaints that the NPT is asymmetrical between nuclear and non-nuclear powers persist. Inspection of nuclear facilities is mandatory for non-nuclear signatories but not for others, for example. But to make things worse, the past few years have provided vivid examples of how relatively unsophisticated countries such as Iraq and North Korea can seriously set out on the road to nuclear power. Plainly but unremarkably, it is now technically easier to make fissile material than to make the simplest means of delivering nuclear weapons in anger; neither Iraq nor North Korea has an aircraft industry.

Calculations

It will be asking a lot of the non-nuclear powers showing up in Geneva in 1995 to put these awkward examples out of mind. (To be fair, the evidence that North Korea is bomb-making is so far circumstantial only; US satellite photographs of clandestine and supposedly nuclear facilities may be otherwise interpreted.) Even stolid non-proliferators may be driven to think twice if they believe neighbouring

states with whom they are on bad terms may share Iraq's ambitions. Thus India and Pakistan, not members of the NPT but still divided over issues such as the status of Kashmir, might logically conclude that a small stock of nuclear weapons is the only means of deterring a sudden attack by the other. The subcontinent is not the only place where people will be making such calculations.

But it helps a little that, since the 1990 review conference, France and, to a lesser extent, China have undertaken to follow the spirit of the NPT and that South Africa, previously insistent that it had no nuclear ambitions, has confessed to having had a small stock of fissile material which has now been dispersed. Indeed, only last week, Argentina acceded to the Treaty of Tlatelolco, which guarantees South America's nuclear-free zone. These developments are not central to the survival of the NPT, but they show that the wind sometimes blows in the right direction.

Negotiation

In the hope of helping it blow more strongly, the nuclear powers responsible for the Partial Test-Ban Treaty (the United States, Britain and Russia, standing in for the former Soviet Union) plan to begin negotiating a comprehensive test-ban treaty in the new year. The hope is that if these three governments can agree to abandon tests, it will be possible to tempt other nuclear powers, of which China is the most important, to join in as well. Ironically, at any other time in the past 25 years, such a deal would have seemed to the non-nuclear powers to fulfil the NPT promise of strategic arms limitation. But what with Iraq, North Korea and cryptonuclear Israel, not to mention the growing uncertainty about the ultimate fate of nuclear weapons in the Ukraine, the non-nuclear members of the NPT may now be unconvinced. One possibility is that 1995 will be the occasion when present members wishing to keep the nuclear option open declare themselves and leave the treaty. Curiously, that may be the best course; then at least the world would know who is toying with making nuclear weapons. It will be worse if the black sheep remain in the treaty to make a mockery of it, as Iraq and probably North Korea have done.

So, in 1995, it will not suffice for the nuclear powers to boast of the quite remarkable array of arms control agreements they have negotiated in the past decade. Nuclear weapons have been made virtually irrelevant to security in Europe by the agreement on intermediate nuclear forces,



strictly strategic missiles in the United States and Russia will be tightly constrained once some US president puts the SALT-II treaty to the Senate (which requires that Russia should emerge from chaos) and there may be a comprehensive test-ban as well. When the whole world is alarmed by the prospect that regional conflicts will be more common now that the Cold War is over, and by the likelihood that some of the participants will have nuclear weapons at their disposal, even the NPT itself may seem an irksome irrelevance to many.

Understanding

So what is to be done? The first need is an understanding between the founding nuclear powers and China, whose influence now is like that of the Soviet Union when the NPT was negotiated. Then, it was easy enough to calculate that the two superpowers would discover a mutual interest in restraining others from holding nuclear weapons while remaining free to improve on their own. China is a less conventional power. Two decades ago, during the Cultural Revolution, there was even Chinese talk of how nuclear weapons in the hands of smaller countries would make the world a more democratic place. Now, in the real world, it is unthinkable that China would benefit from the emergence of North Korea as a nuclear power; it would quickly be emulated by Taiwan and South Korea. Even Japan would be compelled to recast its constitutional detachment from strategic conflict. But it is also unthinkable that North Korea's probable violations of its obligations under the NPT can be dealt with without the consent and even the participation of China. Speculation in the United States that a few cruise missiles directed at the supposedly clandestine sites would do instead is unhelpful, to say the least.

There is an equally urgent need to deal with the nuclear weapons now stranded in the Ukraine as well as those still deployed at Russian launching silos. The Ukraine's legal title to the nuclear weapons on its territory is disputable, but irrelevantly. *De facto* ownership is what matters. And Ukraine does not hide its price for the dismantling of its nuclear stocks: help with the general economic problems that now, after three years of wilful neglect, threaten the kind of chaos that hyperinflation brings. Only the United States and Russia, despite the other things on their minds, can resolve this problem. While doing so, they could aim at a more transparent arrangement for safeguarding their own bilateral agreement on strategic arms than there is at present. That will have to be plain for all to see by 1995.

Rights

There remains the biggest problem of all, that of deciding how the world can live safely when many other states have become nuclear powers. In principle, of course, a decision to make nuclear weapons lies squarely within the sovereign rights of nation states. A government can freely stay outside the NPT, or resign its membership, and decide to equip its forces with nuclear weapons with impunity, just as if it were

deciding on the kinds of blackboards to be used in village schools. In reality, it is not as simple as that: neighbouring states have views on such developments and can express them in all kinds of unfriendly and damaging ways. More powerful and more distant states can also intervene, twisting arms when necessary. But in the absence of an international understanding that making nuclear weapons is wrong, there is nothing that can be legally done.

Yet this is supposed to be the era of what was called, just a few years ago, the "new world order", itself sustained by the success with which the United Nations intervened to roll back the invasion of Kuwait by Iraq. Is it conceivable that the same principles might be made to apply to governments embarking on the manufacture of nuclear weapons? Certainly not, at least if the intention is to bring proliferators to compliance by military intervention, which among other things would be beyond the power of even the most powerful members of the United Nations. But the goal of non-proliferation is not to outlaw nuclear weapons as such, but to prevent nuclear war. How is that to be accomplished in a world in which the weapons themselves are more widely available than in the past?

Danger

That, sadly, is beyond the scope of the conference arranged for 1995, which will by definition exclude many of the governments whose consent to a wider agreement would be necessary, those of South-East Asia, the Indian subcontinent and of Israel, for example. In the circumstances, the 1995 treaty conference is bound to seem a hollow business to its intended participants. The recitation of what the nuclear powers have done in arms control will also be a reminder of the problems remaining outside the scope of the NPT. They will be expected to approve ever-more stringent (and costly) arrangements for safeguarding their own nuclear industries while knowing that declared would-be proliferators remain exempt. There is, in other words, a danger that the NPT will collapse because it seems irrelevant, shutting the stable door after the horse has bolted.

That is why the 1995 meeting should be accompanied by a different kind of negotiation, one aimed at reducing the risks of nuclear war involving smaller nuclear powers rather than elaborating the present NPT. It would, of course, be a messy business, requiring that people should anticipate the causes of different regional conflicts and put in place the safeguards that might help to avoid them. It helps that there has been some progress in previously intractable fields, this year's agreement between Israel and the Palestinians, for example. It should help even more that governments are beginning to appreciate that the time is fast approaching when civil nuclear power will again be a necessary means of generating power, as the restraints of global warming treaties begin to bite. There is, in other words, a carrot as well as a stick to help win general consent to a new and more effective regime for the avoidance of nuclear war. Will the world's governments recognize that in time for 1995? □

Misconduct charges against Gallo withdrawn after Popovic decision

Washington. The Office of Research Integrity (ORI) of the US Department of Health and Human Services last week drew down the curtain on its five-year effort to convict AIDS researcher Robert Gallo of scientific misconduct.

The ORI's action took the form of a terse letter to the appeals board, which has ultimate authority in cases involving the National Institutes of Health. It follows the board's decision the previous week to dismiss charges of misconduct against Gallo's co-researcher, Mikulas Popovic, on the grounds that the evidence produced by ORI did not stand up to legal scrutiny (see *Nature* 366, 99; 1993).

Aware that it was not able to meet such legal standards at a forthcoming hearing against Gallo, ORI told the appeals board that it felt compelled to withdraw its finding that Gallo deliberately included a false sentence in the key paper identifying what is now known as human immunodeficiency virus (HIV) as the cause of AIDS.

Gallo says he is "delighted" with the dismissal, and that he feels "completely vindicated" by the board's verdict. He says that he now intends to "redouble my efforts in the fight against AIDS and cancer".

The paper was published in the 4 May 1984 issue of *Science*. Popovic, who worked in Gallo's laboratory at the US National Cancer Institute and was the first researcher to grow the virus in sufficient quantity to develop a blood test, was the lead author.

More than a year ago, ORI found Popovic guilty on four counts of scientific misconduct, based on statements made in the paper. After a 12-day hearing, however, and an extensive line-by-line analysis, the appeals board exonerated Popovic, saying that it had expected to find "at least a residue of palpable wrongdoing", but had not done so.

Popovic's appeals hearing was the first occasion during which the evidence for and against the charge of misconduct was presented in open court. It was an environment that allowed his attorney, Barbara Mishkin, to cross-examine ORI's witnesses and publicly present evidence on Popovic's behalf.

The appeals board eventually decided that ORI could not prove either that the *Science* paper contained untrue statements, or that it contained "intentional falsifications".

ORI says that it withdrew its case against Gallo because the appeals board has set unreasonably high standards of proof of scientific fraud. Chief among them is the idea, based on legal concepts of evidence, that proof of fraud requires evidence of intent, and that someone has been defrauded.

As the essential data in the *Science* paper

were never challenged, ORI was left to prove that a sentence which, in its view, failed to describe accurately how scientists in Gallo's laboratory used a virus from the Institut Pasteur in Paris was deliberately intended to mislead.

Gallo, who had briefly grown the French virus in his laboratory (but says it did not grow well) wrote: "However, it is possible that this is due to insufficient characteriza-



Robert Gallo

tion of LAV [the French virus] because the virus has not yet been transmitted to a permanently growing cell line for true isolation and therefore has been difficult to obtain in quantity."

ORI said that this sentence was misleading, as Gallo's laboratory notebooks show that LAV had in fact been successfully grown, even though ORI has conceded that the notebooks do not show LAV growing long or vigorously.

Gallo denies ORI's interpretation. He claims that the sentence was intended to indicate that the French researchers had not grown LAV in a permanent cell line, and therefore could not obtain it in quantity.

Gallo says he believed that the virus that did grow well in his own laboratory — identified as HTLV-III — was distinct from LAV.

Several years after the paper was published, tests by Gallo and by Luc Montaigner of Pasteur proved that the disputed virus was a contaminant that originated in France, but that both laboratories were unaware of this at the time of their original studies.

ORI officials have argued that their dismissal of the Gallo case is based on legal technicalities rather than its merits. Interestingly, its critics include AIDS activists from Project Inform in San Francisco.

Martin Delaney, director of Inform, claims that this is an attempt by ORI to save face by suggesting the failure of its case is due to a new definition of scientific fraud. "That is completely untrue", says Delaney, pointing out that the appeals board's explicitly tested the Popovic and Gallo cases against a definition of misconduct adopted by the department in 1989.

ORI is now working on revised guidelines that will allow a scientist to be convicted of fraud only if he or she knew "or should have known" that a sentence in a manuscript might be misleading — an approach that opens up a whole new argument about whether scientists should be legally held to more stringent standards of behaviour than those applied to other academics.

Barbara J. Culliton

French keeps up claims to AIDS patent

Paris. France has repeated its determination to obtain from the United States the credit and the royalties for the discovery of the AIDS virus, following last week's decision by the US Office of Research Integrity to terminate scientific misconduct proceedings against Robert Gallo (see above). "The decision has no bearing on our position," said both the Ministry of Higher Education and the Pasteur Institute in separate statements.

France argues that allegations of scientific misconduct have always been secondary to the wider public policy issues which, it claims, have been raised by the dispute between Gallo and researchers at the Pasteur Institute in Paris over who can claim to have discovered the virus.

French officials claim that everyone — including Gallo — now agrees that the so-called HTLV-III virus used by Gallo to make the US AIDS test was in fact the LAV virus previously isolated by researchers at the Pasteur Institute. (The institute also emphasizes that it supplied Gallo with the LAV virus on the explicit condition that he did

not use it for commercial purposes).

This interpretation first emerged two years ago, following investigations by John Crewdson, a reporter with the *Chicago Tribune*. But the United States has refused to renegotiate the 1987 agreement with France that provides for an equal split of the \$8 million or so annual royalties from the patent for the AIDS blood test. The agreement was based on the understanding that both Gallo and the French had independently discovered two separate strains of the AIDS virus.

This week, Weil Gotshal and Manges, the Washington law firm representing the Pasteur Institute, said that the US Department of Health and Human Services and the National Institutes of Health were bound by "conscience" and their "commitment to scientific integrity" to "bring the 1987 position of the US government into conformance with what everyone now understands to be the truth." However the Clinton administration has previously indicated that it has no intention of re-opening negotiations with France.

Declan Butler

Milk hormone faces new hurdles on way to market

San Francisco. Dairy farmers in the United States, worried that milk may lose its 'pure' image through being associated with genetic engineering, are hesitating before adopting the widespread use of bovine somatotropin (BST), the hormone that can increase yields by up to 25 per cent.

The US Food and Drug Administration announced on 5 November, after nine years of review, that it had decided that bovine somatotropin is safe and effective, and could be sold to dairy farmers. But it is imposing a 90-day moratorium before BST goes on sale, and this together with years of controversy and the continued opposition of groups of farmers, remain a challenge for Monsanto Co., which has developed the hormone treatment for increasing yields.

The hormone has led to demands from consumer groups for genetically engineered foods to be labelled as such, and to declarations by dairy companies that they will not sell milk from cows treated with BST. Activists have promised to organize boycotts and demonstrations against the product.

Surveys have shown that between 15 and 40 per cent of consumers are likely to stop or reduce their milk consumption if farmers begin to use BST. Such surveys have prompted many dairy buyers to decide not to sell BST-generated milk. The California Milk Advisory Board, for example, a marketing coalition for the biggest dairy state in the nation, recommends that dairy farmers

not use the drug.

A number of dairy produce companies have said they will not accept milk from cows treated with BST. Several may label their milk as 'BST-free', even though they admit it will be hard to monitor use of the drug by their suppliers. Three major drug companies promised earlier this year not to use the hormone in their infant formulas.

Monsanto, however, claims that many farmers want its product, according to Tom McDermott, a spokesman for the company. Monsanto plans to use the 90-day period before the hormone goes on sale to convince opinion leaders among animal nutritionists, veterinarians and dairy farmers of its value.

As the drug clears regulatory hurdles, the debate surrounding its use may shift from safety to economics. BST will not actually be sold until February because of the moratorium, set by Congress to allow the White House to study the potential economic impact of the treatment, including its impact on small dairy farms.

Russ Feingold, a Democrat Senator from Wisconsin who was responsible for the bill imposing the moratorium, said he intends to use the study to ask President Clinton to extend the moratorium, to enact a labelling rule, or to tax dairy farmers who use BST. Such a tax would be justified, he said, because of the costs to government of buying up unsold milk.

Sally Lehrman

Protest group turns spotlight on university researcher

Munich. Animal rights activists have placed the name, address and telephone number of a university neuroscientist on a large billboard in the centre of Munich, following the decision of a newly-appointed ethics committee to overturn the ruling of an earlier committee not to approve experiments that she had proposed on primates.

The animal rights group also distributed leaflets in the researcher's home village describing scientists who use animals in their research as criminals who should not remain anonymous. Police ordered the removal of a black cross — a sign of death — which had been painted over the billboard.

The researcher involved is Jean-Alice Büttner-Ennever, a professor of neuropathology at Ludwig Maximilian University. She has since received many death threats. A spokeswoman for the animal rights group, Tierversuchsgegner München, said that the group would go as far as it could, short of actually harming people, to achieve its goal of abolishing the use of animals in biomedical research.

Representatives from the university, its physiological institute and its neurological clinic have all come to Büttner-Ennever's defence. Further support has come from patients who have benefitted from her research, which has led to a method of identifying nerve pathways responsible for balance and eye movement disturbances following stroke.

Her research is supported as a 12-year 'special research programme' by the Deutsche Forschungsgemeinschaft. Last summer, an ethics commission reversed previous decisions that had been regularly taken to approve her work. But this decision has now itself been overturned.

The animal rights group says that singling out scientists for a personal campaign is an effective form of protest. "We want to give [animal researchers] sleepless nights", says the spokeswoman.

The incident is one of several in Germany in the past few years, even though the number of animals used in medical research has fallen. In 1989, 2.64 million animals were used in the old *Länder* alone; in contrast, the total number used last year in the whole of Germany was 2.08 million.

Meanwhile, research scientists with the British pharmaceutical company Glaxo are also being targeted by an animal rights group. The Animal Rights Militia has claimed responsibility for a series of incidents designed to dissuade researchers from joining the company's new research complex at Stevenage, including sending a hoax bomb to the home of one employee.

Alison Abbott

UK report criticizes plan for PhDs

London. Proposals by the British government to introduce a mandatory masters qualification as part of the PhD degree were criticized in a report issued this week by the National Commission on Education, a body set up two years ago at the suggestion of Sir Claus Moser, then president of the British Association for the Advancement of Science (BA).

The proposal was made in the recent white paper on science, published in May. But the commission says that, since the government does not intend to provide any more money for the revised research degrees, the number of students taking doctoral courses will inevitably fall. This will increase the difficulty UK universities already face in employing researchers at rates competitive with industry, or with universities in other countries.

Recommendations made by the commission include increasing the number of graduate schools in universities, raising the morale of academic researchers by improving pay and career structures, and

providing training skills to university lecturers, most of whom have no formal training in teaching.

The commission also calls for a proper financial strategy for future expansion, worked out between the government and the higher-education institutions. It says that the money to support such expansion cannot continue to come primarily from gains in efficiency, as it has done over the past decade or so; it suggests that its recommendations are funded through greater contributions from both industry and students.

The recommendations are part of the commission's two-year, independent review of education and training in Britain. The study was initiated by the BA in the light of serious shortcomings in education and training standards, particularly in comparison with other European countries, and the refusal of the government to set up its own study.

Maxine Clarke
* *Learning to Succeed* National Commission on Education, Heinemann, £4.99

Foundation seeks enhanced role in Europe

Strasbourg. The European Science Foundation (ESF) wants to change direction and establish itself as an influential European science policy forum capable of playing a key role in shaping European research. Next week's annual assembly of the foundation will vote on a package of changes designed to raise its profile and influence at both national and European levels.

The ESF was set up 19 years ago, and since then has been quietly but effectively running scientific networks, workshops and a limited science programme. For the past year, however, it has been engaged in defining a more important role for itself.

One thing will not change. The foundation, which at present has 54 member organizations from 20 different countries, will continue as a non-governmental, pan-European organization concerned with all fields of fundamental research.

But the opening up of eastern Europe and the increasing importance of the research programme run by the European Commission — much of which is now directed to fundamental research and overlaps with ESF activities — have prompted a reassessment of its general aims.

Pressure for change in Europe has also come from national research organizations. In the past year, a group of heads of such organizations, known as the 'Eurohorcs', have been seeking more direct control over the research money spent by the European Community (EC). They would like the ESF to become a much stronger forum through which their voices could be heard in Brussels.

Peter Fricker, a founder member of ESF and former chief executive of the Swiss National Science Foundation, who took over as secretary general in April, is keen for the foundation to develop in this direction. He believes it would allow the foundation to act as a general champion of fundamental science, which he feels to be under threat in many European countries.

Next week's assembly of the ESF will debate proposals to place a financial limit on ESF science programmes, at least to their present level, while extending its strategic roles, and to develop stronger and more formal links with both the EC commission and the countries of central and eastern Europe.

ESF officials would also like the foundation to play a central role in defining future requirements for major scientific facilities in Europe, in assessing long-term prospects for scientific developments, and in monitor-

ing the scientific output of, for example, major national and European laboratories.

The proposal to adopt a more strategic role is likely to be accepted by the meeting. But some controversy may also arise. For example, the ESF will have to ensure that it does not duplicate the activities of other organizations, such as the Organization for Economic Cooperation and Development.

Fricker also recognizes that the foundation cannot force the EC to accept its proposals. While the commission is likely to react favourably to a system which offers structured advice from the basic science community, it cannot legally delegate any decision-making to an outside body.

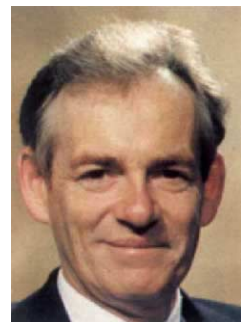
Fricker says he is aware that issues such as the delegation of activities such as grant refereeing — an idea that the ESF has considered in the past — could raise legal difficulties. Similar direct approaches to the commission by the Eurohorcs have not been wholly welcomed, since they are seen as a threat to the authority of the commission.

More acceptable to the commission is likely to be an extension of *ad hoc* cooperation with the ESF. There have already been some successful examples of this, such as



Peter Fricker

Sir Dai Rees, secretary of Britain's Medical Research Council, is expected to be elected as the new president of the ESF at its general assembly next week. Rees will be appointed for a three-year term. He will take over from Umberto Colombo, who will not be standing for re-election following his appointment as Italy's minister of research and universities.



the European Research Conferences, which are modelled on the US Gordon Conferences and are funded by the EC and organized by the ESF.

Fricker says that the ESF is likely to rely much more in future on member organizations, for example to manage specific scientific activities. How the changes within the ESF might be paid for is at present being left open until its future plans become more concrete.

Alison Abbott

Kohl limits money for universities

Munich. Helmut Kohl, the German chancellor, promised at a long-awaited education summit in Bonn last week that he would set up a committee to look at ways in which the country's federal higher education law might be changed to give universities more freedom to set entrance examinations and control their own budgets.

The summit was originally called by Kohl in spring of last year in response to charges that the universities face too many students staying for too long.

But Kohl made no commitment to increase funds for building and maintenance of higher education institutions, and announced only a small increase in funds for research in the new *Länder*. Many regional representatives, who want the federal government to pledge more cash to help reforms, were disappointed by his speech.

The prime ministers of Germany's 16 *Länder*, which have full responsibility for running institutes of higher education in their regions but depend on federal funds for building, maintenance and large equipment, had asked the government to increase the 1994 building budget from DM1.68 billion (US\$950 million) to DM2 billion.

This in itself was a compromise. The Wissenschaftsrat — Germany's science advisory council — has calculated that at least DM2.3 billion is required to maintain exist-

ing buildings at universities and other higher education institutes, even without any expansion. But Kohl refused any increase.

In sticking to its guns, the federal government runs the risk that the prime ministers, who control the Bundesrat, Germany's upper house, will delay the passing of the whole federal 1994 budget later this year by taking the unusual step of referring the issue to the Vermittlungsausschuss, a mediation committee with the Bundestag.

The Wissenschaftsrat had also recommended an increase in federal spending of DM300 million on research. But, in light of the government's attempts to keep a tight lid on public spending, Kohl has agreed to only half of this increase, and has also earmarked the extra money for industrial research in the new *Länder*.

Gerhard Neuweiler, head of the Wissenschaftsrat, says that the government appears to have no interest in investing in universities. He suggests this was the main reason for the summit being delayed for more than a year, and finally taking the form of a series of position statements which were not open to discussion.

Some see Kohl's promise to set up a committee to look at future reforms as a delaying tactic. But others are more optimistic, and believe that the law could still be reformed next year.

Alison Abbott

Danish council fails to satisfy its critics

Paris. Danish scientists are claiming that the recently-created Danish National Research Foundation, set up with the proceeds of the privatization of a number of publicly-owned insurance companies, has spent its first round of five-year grants in ways that fall far short of its goal of "funding unique Danish research at the international level."

Two years ago, the government decision to create the research council with the DKr 2 billion (US\$297 million) proceeds of the privatization raised protests from scientists, who described it as a "cigar box", the Danish term for political earmarking (see *Nature* **360**, 517; 1992). They argued that the money would be better spent through the country's existing five research councils. The government apparently wanted to avoid spreading the money too thinly, and to control how it was spent.

The foundation has set out to provide substantial grants to research groups which are of international standing and capable of managing a large research effort. To qualify, the research proposed must not be well supported through existing channels, and must also open up new possibilities for international collaboration.

In addition, research supported by the foundation must be "relevant to Denmark from a cultural point of view, or because of demands from the public or private sector."

Whatever their feelings about the goals of the foundation, many Danish researchers have expressed surprise at some of the 23 research groups chosen by the foundation to share DKr750 million of funding over the next five years. "Everyone here was extremely disturbed and alarmed about the distribution" says one senior researcher on the medical faculty of the University of Copenhagen, claiming for example that

many of the winners have had a poor publication record, lack experience, and are only "loosely connected to the international scientific community".

The foundation's nine-member board short-listed 50 applications from the 350 it received. Peder Olsen Larsen, the Aarhus chemist who is the foundation's first chairman, says it was impossible to send all 350 applications out to external review (the referees were paid). "You don't need to be an expert to see if people are international, or if their research plan is coherent," says Larsen. The fifty short-listed proposals were sent out to foreign referees; but the board had the final say in who received grants.

The board was appointed in 1991 by Bertel Heerder, the then research minister. He nominated Larsen as chairman, and Ulrik Lassen, senior vice-president of Novo Nordisk, as deputy chairman. Heerder also chose seven other board members from candidates proposed by the research councils and other scientific bodies.

Jens Rehfeld, an endocrinologist at Copenhagen University Hospital who had his application to create a peptide centre rejected despite very favourable reports from foreign referees, criticizes the fact that the board's members are no longer active scientists. Rehfeld also argues that nine individuals cannot have sufficient expertise to cover all fields of science and the humanities.

One individual close to the board agrees with both of Rehfeld's assertions. But he says that the referees provided the necessary scientific expertise. And a board member, answering the charge that not all the groups selected for support already enjoy an international reputation, admits that in some cases the foundation selected "groups who are not yet strong, but who we expect will be

given this money". Rehfeld claims, however, that the board superimposed additional criteria, which he describes as political, on the referees' assessments.

Larsen defends the choice of winners, and points out that these were agreed unanimously by the whole board. Board members — and some other scientists — also point out that controversy is to be expected whenever large amounts of money are distributed, especially when decisions have to be made on projects covering all academic disciplines.

How well the foundation has spent what represents a significant share of the Danish research budget will become known in 1996, when external reviewers evaluate the selected projects, and recommend if they should continue to be funded. Meanwhile falling interest rates threaten to reduce the return on the foundation's capital, and to leave it with little money to distribute in its next funding round, due in 1998.

Declan Butler

UK rules out new gene therapy laws

London. Britain has for the time being ruled out new legislation regulating the field of gene therapy. Tom Sackville, the Parliamentary Secretary for Health, told a meeting in London on Monday organized by the Bioindustry Association that the British government felt a legislative approach to an evolving field "would not be appropriate".

Sackville also used the opportunity to announce the sixteen members of a new Genetic Therapy Advisory Committee. This committee will replace the Committee on the Ethics of Gene Therapy, which has been chaired by Sir Cecil Clothier. It will be headed by Dame June Lloyd, formerly professor of child health at the Institute of Child Health, London.

The new committee will review all proposals for gene therapy research on human subjects in Britain, and will advise Mrs Virginia Bottomley, the health minister, on technical developments in gene therapy and their implications. Its members range from geneticist Martin Bobrow and theologian Keith Denison to the broadcaster and journalist Nick Ross, and Brian Richards, chairman of British Bio-technology Ltd.

Bob Williamson, professor of biochemistry at St Mary's Hospital Medical School in London, who is closely involved in gene therapy trials of a potential treatment for cystic fibrosis, says that the latest announcement should act as a stimulus for European biotechnology companies. Unlike their counterparts in the United States, he said, European companies had not "gripped the nettle" of gene therapy.

John Hodgson

CNRS moves home — and cuts down on staff

Paris. France's Centre National de la Recherche Scientifique (CNRS) has opened a new headquarters at the Michel-Ange campus in southeast Paris. The organization, which is France's main source of funds for basic research, says that the move from the centre of Paris to the new site, which comprises five modern buildings and a mansion built in 1710, symbolizes a shift in the priority of its administration from day-to-day management to a greater concern with steering policy. To this end, the new site brings together the administration's seven scientific departments, its national committee, and Institutes of Sciences of the Universe, and of Nuclear and



Particle Physics, which were previously scattered between seven sites around Paris. At the same time, the CNRS has trimmed the number of administrators from 1,100 to 700.

D.B.

Australia backs off CSIRO merger with nuclear body

Sydney. Members of the governing board of Australia's Nuclear Science and Technology Organisation (ANSTO), the body which runs two research reactors outside Sydney, have agreed to resign following the failure of an attempt to merge it with the Commonwealth Scientific and Industrial Research Organisation (CSIRO).

Their resignations will allow for the implementation of a compromise arrangement under which ANSTO's board will be reconstituted to include a number of CSIRO nominees. The move is being seen as a way of encouraging closer collaboration between the two organizations.

The original suggestion of a full merger between the two had been proposed to the government by Chris Schacht, the minister for science and small business, as a way of achieving this goal and also saving money. However this proposal was rejected by the Federal cabinet of ministers. Some members are thought to have been unconvinced by Schacht's argument that the merger of the two agencies would have saved Aus\$25 million (US\$38 million) a year.

Defending the government's decision, Schacht argued that Australia's Greens and other minority parties, who hold a significant voting bloc in the Senate, would have delayed the merger legislation because of general opposition to all nuclear activities.

Behind the decision to reconstitute the ANSTO board lie reports of conflict between Schacht and the board, which will continue to function in a 'caretaker' role until Christmas, and is headed by Richard Collins, professor of applied physics at Sydney University.

Schacht is reported to have accused the board of bad management. Such charges have been fanned, for example, by recent problems with the installation of a supercomputer, the subject of an audit report soon to be released by the government.

ANSTO and its supporters, however, argue that occasional mistakes are inevitable when an organization undergoes large-scale change, as the nuclear body has been doing over the past few years. They also counter Schacht's charges with accusations of excessive ministerial interference in the management of the organization.

Schacht says that the new ANSTO board will review the operations of the organization to ensure that the government's objectives — including developing closer links with CSIRO — are being met. The arrangements will be reviewed in two years, says Schacht, warning that unless the government's objectives are achieved, further consideration will be given to merging the two organizations.

Mark Lawson

Spanish physicists are spoilt for choice of accelerators

Barcelona. The Spanish government has opened discussions with three regional governments over how much support it is prepared to provide for their separate plans to build particle accelerators. But the plans have been criticized as being insufficiently coordinated, driven by political motives, and a drain on scarce financial resources.

The respective plans are for a large synchrotron in Barcelona, the capital of the northern autonomous region of Catalonia; a smaller synchrotron in Seville, in the southern region of Andalusia; and for a third accelerator — possibly linear — in Madrid. Each has sprung from a desire by the respective regions to attract the biggest slice of the national science budget.

At the official opening of Cartuja '93, Spain's new science park occupying the

Typical! You wait ages for a new particle accelerator, then three come at once...



groups of Expo '92 in Seville, Elias Fereres, state secretary for universities and research, defended the government's decision to back all three projects. "We want to guarantee access to all Spanish researchers interested in using these machines," he said.

Originally, Spain had hoped to build a tau-charm factory. However these hopes crashed last December when the European particle physics laboratory, CERN, refused to support the proposal. Fereres reached agreement in principle with Seville, Madrid and Barcelona to support instead a synchrotron for Barcelona and a cyclotron for Seville, principally for medical purposes.

The news of plans for a synchrotron in Andalusia, instead of the proposed cyclotron, has therefore taken scientists by surprise. Manuel Lozano, head of nuclear physics at the University of Seville, says that it would be a "disaster" if Catalonia and Andalusia ended up with the same sort of machine, which could not be justified in such a small scientific community.

"No-one knows where the terms got changed", says Lozano. The proposed machine would be housed at the Cartuja site, and is expected to cost up to Pts1 billion (\$7.5 million).

The Catalan synchrotron, already under construction in Barcelona, is estimated to cost Pts11 billion. The 2 GeV machine will probably be housed in a technology park near the city's Autonomous University, and has already encountered criticism.

First, the number of potential users may be too small to make it competitive. By the time the synchrotron is complete at the beginning of the next century, Geneva's 6 GeV European Synchrotron Radiation Facility will be in full operation. So too will several other smaller accelerators.

Second, the money Catalonia is prepared to pay for the synchrotron is more than the total sum the region has put into science and research since the death of Franco. Regional government officials justify this expenditure as stimulating industry and research. But many scientists are critical of this choice of priorities.

Meanwhile in Madrid, the autonomous government has rejected a previous plan to buy a smaller (0.7 GeV), second-hand synchrotron in favour of a small linear accelerator, housed in the University of Madrid.

Last month, talks opened between central government and the local governments of Andalusia and Catalonia on how budgets for all of these accelerators should be shared between them. According to Fereres, funding will come from the central government through the national science plan. But none of the plan's agreed budget has been earmarked for this purpose.

The country already has one accelerator, which is not yet in operation, and whose schedule has also been disrupted by politics. In 1989, Lozano raised Pts150 million from the European Commission to buy a small ion electrostatic tandem accelerator.

But Luis Oró, general secretary of the national science plan, persuaded Lozano to buy a bigger machine (4 million volts) for national use. The additional Pts350 million required for this machine has also been provided by the commission.

The accelerator was built in the United States, and is ready for delivery. But it remains in storage. The central and regional governments have now decided that it should be housed at the Cartuja site, which is having difficulty in attracting permanent occupants, rather than in the grounds of the university, as originally planned. But the Cartuja building will not be ready for the accelerator until at least 1996.

Luis Angel Fernández Hermana

Japan debates proposal for a 'super-ministry' of science

Tokyo. An audacious plan has been put to the Japanese government for re-grouping all the science-related sections of its ministries and agencies into a single ministry of education, science and culture, a move that would have profound consequences for the administration of government-funded science.

The plan is one of a series of proposals for government reform put to the new administration of prime minister Morihiro Hosokawa. In particular, it is being suggested that the number of ministries and agencies be reduced from over 20 to six.

At first glance, the proposals seem too radical to be taken seriously. But they come from the powerful *ad hoc* committee on administrative reform, the body which in the early 1980s recommended the privatization of the national railways, Nippon Telegraph and Telephone (NTT) and Japan Tobacco.

The latter goals were each realized, thanks in large part to the strong leadership of the then prime minister, Yasuhiro Nakasone. In the same way, the fate of this latest proposal, which will be fiercely resisted by the ministries themselves, depends very much on Hosokawa and his coalition administration.

The committee that has made the proposal is chaired by Eiji Suzuki, a senior adviser to Mitsubishi Kasei chemical company, and has two subcommittees made up of academics, industrialists and journalists. These discussed the possible reform of the central government for three years before presenting their recommendations to Hosokawa at the end of last month.

At present, the Ministry of International Trade and Industry (MITI), the Ministry of Education, Science and Culture (Monbusho in Japanese) and the Science and Technology Agency (STA) are the three key players in Japan's government science. However, the Ministry of Health and Welfare, the Ministry of Agriculture, Forestry and Fisheries, the Ministry of Posts and Telecommunications and several other small agencies also make significant contributions to government research.

The committee's overall goal is to increase the efficiency of central government by eliminating unnecessary duplication of effort and breaking down the bureaucratic walls that separate ministries and agencies with similar or overlapping interests.

Thus their report calls for the amalgama-

tion of the present 22 ministries and agencies into six ministries, covering areas such as "lifestyle" (including health, pensions, housing and the environment), "industry", and "national land" (including roads, transport, and communications).

One such body would be a giant ministry of education, science and culture (not to be confused with the present Monbusho). Many government officials are horrified by such a suggestion. Officials of the STA, in particular, are opposed to the idea of working in a ministry where they might be assigned to work on education in primary and high schools, rather than science and technology. (Monbusho's preoccupation with schools is one of the great weaknesses of the university research system that the ministry oversees).

But MITI officials are more receptive to the calls for change. Reform-minded government officials in MITI have for several years been calling for more coordinated interaction between Monbusho, MITI and STA in developing and carrying out science and technology policy. They think the debate opened up by the Suzuki committee will be very healthy, although they do not expect the recommendations to be implemented as proposed.

Much depends on the future of the Hosokawa administration. At present, Hosokawa is preoccupied with pushing through the reform of Japan's election system on which he has staked his future as prime minister. If he succeeds, the next target will be to loosen governmental controls on nearly all aspects of Japanese life.

Behind these moves is a fierce power-struggle between the bureaucracy and the Hosokawa administration. Members of the new administration, in particular Ichiro Ozawa, a former LDP politician, want to reduce the power of government officials. Ozawa, who co-chairs Shinseito, one of the most powerful parties in the eight-party coalition, is proposing that senior officials should no longer participate in Diet committees, and that politicians should instead be appointed in the ministries and agencies to perform this role.

Such moves, however, are being strongly resisted by government officials, who argue that politicians must first clean up the corruption in their ranks before playing a stronger role in the administration of policy.

David Swinbanks

Opposition fails to dent commitment to plutonium

Tokyo. Japan seems determined to defy international opinion by pursuing plans to use large vast amounts of plutonium in its nuclear power industry, both despite and because of delays in the country's development of fast breeder reactors.

The latest annual report of the Atomic Energy Commission (AEC), released last week, reaffirms Japan's commitment to a policy that will result in Japan increasing its supplies of plutonium from a few tonnes at present to 85 tonnes by 2010. This will be done through both local production and re-imports of reprocessed nuclear fuel.

The United States has long abandoned most of its fast breeder reactor programme, and, under a new nonproliferation policy announced in September (and reinforced in a statement to Congress last week), President Bill Clinton has declared that he does not support the civilian use of plutonium.

Britain and Germany have frozen their fast reactor programmes. And France's Superphénix reactor has been plagued with problems. But the AEC report insists that Japan must pursue the production of plutonium through nuclear fuel recycling, and promote the use of the fuel in conventional and fast breeder reactors.

Japan faced severe international criticism in January when it brought back a tonne of plutonium from a reprocessing facility in France by sea.

Despite the furore, Japan is now planning to ship about another 54 tonnes of plutonium derived from reprocessed Japanese nuclear fuel from Britain and France over the next two decades. And a further 30 tons will be produced by a reprocessing facility under construction in Rokkasho village in northern Japan.

The first few tonnes will be used in the prototype fast breeder reactor Monju, which is expected to reach criticality in April next year. But the supply of plutonium will quickly exceed the needs of Monju. Instead, plutonium will be mixed with uranium fuel for use in ordinary light water reactors.

Given the current low price and ample supplies of conventional uranium fuel, it is uneconomical to use mixed uranium-plutonium fuel, but the decision is being forced on Japan's nuclear power industry because of delays in the development of commercial fast breeder reactors — economic doubts have pushed back the start-up of a commercial reactor until at least 2020.

Meanwhile the plutonium has had to be put to some use to allay concern that Japan is stockpiling it for production of nuclear weapons; hence the decision to burn up plutonium in ordinary light water reactors.

David Swinbanks

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Will Hosokawa force reforms?

UK brain-drain 'no worse'

London. Britain's brain-drain is not improving. But neither has it been getting much worse, according to a report published in London last week which will provide ammunition both for the government and its critics in the debate over the number of British scientists leaving to work overseas.

The report has already been welcomed by William Waldegrave, the minister for science, who has homed in on its conclusions that there has been a slight fall in the number of scientists emigrating, and a small increase in the number coming to Britain from overseas. In each case, the number remains small compared to the total number of scientists in the United Kingdom.

But those responsible for the report have issued a warning about an excessively complacent interpretation of its conclusion. They point out that it still tends to be the highest quality staff who take posts overseas, confirmed by the growing number of fellows of the society with foreign addresses.

The report has been produced by the Science and Engineering Policy Studies Unit (SEPSU) of the Royal Society. It is based on replies to a questionnaire sent to university science departments, asking for details of individual scientists known to have left the country between 1984 and 1992. Five subject areas were covered: biochemistry, chemistry, Earth sciences, electrical engineering and physics.

The report follows up an earlier version, published in 1987 and covering the period 1975–85, which was the first significant attempt to quantify a phenomenon whose existence tends to be known more through personal anecdote than hard data.

Like the earlier report, the new one points out that, at least in terms of overall figures, the net loss of scientists to Britain — in particular, the tendency of postgraduates to seek research posts overseas — is virtually balanced by the number coming to work in the country from abroad.

The numbers involved have changed little. The new report, prepared with support from the Nuffield Foundation, found, for example, that the number of recently qualified postgraduates leaving to work overseas had risen slightly, from 13.4 to 13.5 per cent. The number of more senior staff leaving in the second period was 2.1 per cent, and heads of department 0.3 to 0.5 per cent.

The biggest change was in chemistry, where the number of departing PhDs rose from 12.0 to 16.2 per cent; in contrast, their peers in Earth sciences fell from 23.6 to 12.7 per cent, largely reflecting declining demand in the oil industry. In biochemistry, the emigration rate for new PhDs remained at about 19.0 per cent.

Nor has there been a significant change in the motivations to leave Britain — or for returning to it. Most scientists left for professional reasons, the most widely quoted grounds being enhanced career prospects and "a desire to widen experience"; 73 per cent of those coming back quoted unspecified personal reasons as a motivation, and 64 per cent a desire to return to British culture.

One novel finding of the report is the different statistics between men and women. The study found that, when compared to the total numbers of each category, women were about half as likely to emigrate as men — but equally likely to come to Britain.

The biggest change since the mid-1980s has been the destination of those leaving Britain. The proportion taking up research posts elsewhere in Europe has risen from 22 to 31 per cent. In contrast, there has been an almost identical drop, from 63 to 53 per cent, in the proportion of emigrant scientists taking up positions in the United States.

The government has, perhaps inevitably, taken comfort from the fact that the brain-drain does not seem to be getting any worse. Indeed, the report points out explicitly that those entering Britain's science and engineering base "seem to have been no more likely to emigrate during the past decade than they were during the previous decade."

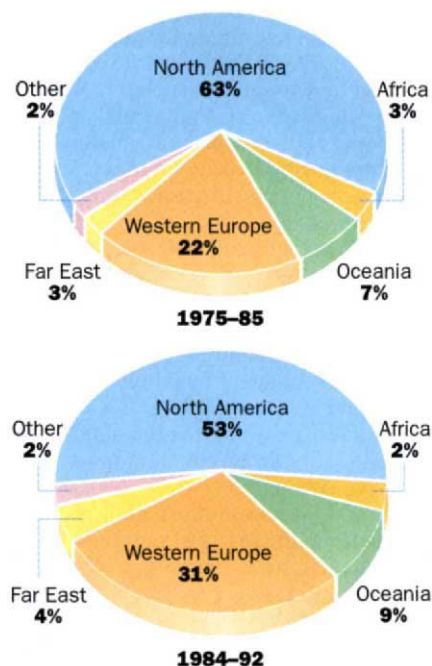
But the report itself warns of the dangers of complacency. It points out for example that those leaving Britain tend to take up long-term positions abroad, returning only for personal reasons. In contrast, those arriving take short-term posts.

Smith also urged the government to take a new look at proposals first put forward two years ago to create so-called Faraday Centres, designed to bring together academic and industrial scientists.

These suggestions were due to be made to a meeting at Imperial College, London, held by the organization Save British Science. According to an advance text of his speech, Smith planned to use the occasion to criticize the impact on science of three beliefs held by the conservative government: the virtues of the market; that the public sector tends to do things less efficiently than the private sector; and the political need for massive cuts in public expenditure.

While adding his voice to the many who have attacked the short-term outlook of both British industry and financiers, Smith claimed that the government has yet to take

Destination of scientists emigrating from the U.K.



"We should be careful not to draw too much comfort from the simple numerical head count" says Ian Nussey, chairman of SEPSU's management board, pointing to the growing number of Royal Society fellows with addresses outside Britain. Almost two-thirds of those replying to the survey felt that emigration was having an adverse effect on British science, and half of these felt that the effect was "serious".

David Dickson

Labour leader pledges support for LHC

London. John Smith, the leader of Britain's opposition Labour Party was expected to confirm this week that, if elected, his party would create a single Ministry for Science, and to pledge support for the proposed Large Hadron Collider (LHC) at CERN.

Smith also urged the government to take a new look at proposals first put forward two years ago to create so-called Faraday Centres, designed to bring together academic and industrial scientists.

These suggestions were due to be made to a meeting at Imperial College, London, held by the organization Save British Science. According to an advance text of his speech, Smith planned to use the occasion to criticize the impact on science of three beliefs held by the conservative government: the virtues of the market; that the public sector tends to do things less efficiently than the private sector; and the political need for massive cuts in public expenditure.

While adding his voice to the many who have attacked the short-term outlook of both British industry and financiers, Smith claimed that the government has yet to take

such criticism fully on board. In particular, its approach to "technology foresight" appeared to be repeating the mistake of demanding "too rapid a pay-back".

The opposition is also keen to bring basic and applied research closer together, and would achieve this by creating "overlapping networks" in which scientists and members of the business and financial community would meet to exchange ideas. One way of achieving this would be through the so-called Faraday Centres; the question of how the academic and industrial research communities "has still not been properly addressed", Smith said.

More resources for direct government funding of research and development, and for the science base, must both be priorities, said Smith. Expecting doctoral students to live on as little as £4,720 a year, for example, is "little short of scandalous".

And he urged the government to increase its support for international projects such as the LHC. "If the British government wants Europe to lead the world, here is a project which it should be actively supporting." □

Making the desert bloom

SIR — Your witty discussion¹ of the study by Nishimori and Ouchi² of the dynamics of wind-blown sand particles treats the subject as of purely intellectual interest with some application as a diversion from boredom during seaside holidays. In fact there may be practical applications important for humanity.

I was an agricultural and building maintenance worker at Atzmona in the Northern Sinai until Israel relinquished the area to Egypt in 1982. The sand, which constantly came from the north (from the direction of the sea) would build up on the north side of any building and of most plants, eventually burying them unless we shovelled it away, a highly labour-intensive occupation.

Sandy desert areas are both a resource for resettling the world's homeless and, as we Israelis have proved, places of tremendous agricultural potential. The heat and the year-round sunshine, together with the quick growth of root systems in the soft sand, give us some of the shortest times from seed to harvest in the Northern Hemisphere, especially because we grow under transparent plastic and use drip irrigation. Musk melons from Atzmona, for example, were always the first in the European markets in the spring. And every season of the year was good for some crop.

The sands, however, were always a problem. We tried to protect fields with plastic or jute screening but the sand quickly built up on the north side of the screens until it topped and covered them.

Soil that we enriched organically by ploughing in harvest waste and root systems at the end of each growing season was also quickly covered by the sands.

Research such as that by Nishimori and Ouchi may eventually find solutions to these problems, making the sandy deserts at least partial solutions to much of the world's hunger and overcrowding. Among the questions such research might answer are the following:

- (1) How might one design more efficient barriers, or devices to deflect moving sand around buildings, roads and fields?
- (2) Although the dunes cover most domesticated plants, date palms and some wild trees and grasses seem to remain uncovered, and the Bedouin are even able to pasture small flocks. What is it about these plants that deflects sand particles?
- (3) Where does the sand come from? Informal observation suggests that, in the Northern Sinai at least, it moves mostly southward, so one may hypothesize that

much of it comes from the sea. What are the dynamics of this cargo-transfer and how would such knowledge make it possible to devise a method to return the sand to the sea or to keep it there?

(4) If human control of wind-blown sand were significantly to alter major geological processes, would the long-term ecological consequences all be salutary?

Frank J. Leavitt

*The Lord Immanuel Jakobovits Center
for Jewish Medical Ethics,
Ben Gurion University of the Negev,
Beer-Sheva,
Israel*

1. *Nature* **364**, 385 (1993).

2. *Phys. Rev. Lett.* **71**, 197 (1993).

Aspirin on trial

SIR — Further to your News item about the use of aspirin in trials as a treatment for HIV (*Nature* **364**, 369; 1993), we have recently reported at a meeting that sulfasalazine, a salicylate compound like aspirin, increased T4 cells in three AIDS patients (*J. Rheumat.*, in the press). The manuscript, now including four patients, has been submitted for publication.

These observations are preliminary, but were made in humans and may be more relevant than the *in vitro* preliminary observations quoted in your article.

E. Disla

H. R. Rhim

A. Reddy

A. Taranta

*Cabrini Medical Center,
227 East 19th Street,
New York, New York 10003, USA*

UNESCO and copyright

SIR — Contrary to your assertion (*Nature* **363**, 382; 1993), UNESCO has been actively concerned with the question of copyright in electronic texts.

The question was first studied within the framework of the joint action between UNESCO and the World Intellectual Property Organisation (WIPO), another UN agency, to find a solution to the copyright problems raised by the use of computers for access (storage and retrieval) of protected works. The studies prepared by the two organizations were first examined in 1979 by a working group of independent experts. Its conclusions were examined in the same year by the Intergovernmental Copyright Committee and the Executive Committee of the Berne Union, which recommended a further

study of the problem at intergovernmental level. Following this recommendation, UNESCO and WIPO convened a Committee of Governmental Experts in 1980 and then a second one in 1982. The latter committee adopted a set of Recommendations to Member States on how to settle the problem.

Since then, there have been rapid developments in technology and telecommunications so that, potentially, information can be digitized, copied, stored, manipulated and delivered electronically anywhere in the world. This has raised new copyright problems for the copyright owners and their representatives as well as for librarians and their users. UNESCO is aware of all these problems and has been working closely with WIPO to establish some guiding principles. Many problems still remain that will be solved only by international organizations working in harmony.

This, I believe, is one further reason why the UK government should rejoin UNESCO; much better to be inside participating than outside jeering.

Ross Shlmon

*Library Association,
7 Ridgmount Street,
London WC1E 7AE, UK*

Russian industry

SIR — In your leading article "Is there a future for Russian science?" (*Nature* **365**, 475; 1993) you make no mention of the need for links between science and industry. In my discussions in Moscow with senior biotechnologists and administrators in June this year, I was impressed by the lack of interest in the application of their science to a rejuvenation of their drug industry. This was particularly poignant since *Realising our Potential* had just been published as a White Paper (policy document) by the British government.

In terms of application, Russian scientists concentrate their efforts on the development of sophisticated drugs and vaccines which they aim to sell for hard currency. When I suggest that they should devote their efforts to their own industry they tell me that it is hopeless in the absence of large amounts of capital for new plant. But I ask myself what life would be like for our research scientists in the life sciences if we had virtually no drug industry. While I agree that it is important to try and ensure that basic science survives in Russia, more effort should be devoted to channelling scientists towards a rejuvenation of their drug industry and health services.

Peter N. Campbell

*Department of Biochemistry
and Molecular Biology,
University College London,
London WC1E 6BT, UK*

Letters submitted for Correspondence should be typed, double-spaced, on one side of the paper only.

Aircraft hazard from volcanoes

SIR — Tilling and Lipman¹, in discussing hazards associated with volcanic eruptions, identified action required to reduce the risk to passenger aircraft. Although it is true that technological advances may not provide a panacea for all risks associated with volcanoes, we believe that the risks to jet aircraft that encounter volcanic ash clouds can be reduced considerably by using appropriate on-board technology, an option not discussed by Tilling and Lipman. We have built and tested² an airborne infrared device capable of detecting volcanic ash clouds from commercial jet aircraft up to 10 minutes before impact. Airborne trials of the device at the Sakurajima volcano in Japan confirm that this seems a promising approach.

Existing warning systems for aviation rely on a combination of ground-based observers, Earth-observing satellites and a tortuous communications network. Our device would be mounted on board the aircraft, providing instantaneous information to the people who need to know — the pilot and crew. Moreover, the device uses passive infrared radiation and works equally well during the day and night. (The Total Ozone Mapping Spectrometer cannot detect volcanic clouds at night, yet several documented encounters with clouds have occurred at night³.)

Airborne volcanic substances cause engines to stall; they clog the pitot-static inlets, oil and fuel filters; they abrade the leading edges of the aircraft; they sandblast the cockpit windscreen; and, because of the sulphuric acid content, they cause corrosion to the skin of the aircraft³. In many documented cases, the hazard was encountered within 1–3 hours of a volcanic eruption. We believe that with the current and planned suite of Earth-observing satellites coupled with the inevitable problems in communication systems, there will be significant delays in pinpointing hazardous volcanic clouds and relaying the information to aircraft in flight. Our on-board device could provide the solution.

Fred Prata

Ian Barton

*CSIRO Division of Atmospheric Research,
Aspendale, Victoria, Australia*

Jeff Kingwell

CSIRO Office of Space

*Science and Applications,
Canberra, Australia*

Colour coded

SIR — With the coming of cheaper colour printing, *Nature* is producing diagrams, such as that showing the species richness of birds in Idaho¹, in colour. Yet possibly 4 per cent of readers are, like me, colour blind, and find many of the colours indistinguishable, particularly reds, greens and browns.

A way of overcoming this difficulty is to use a large field magnifier. Such magnifiers are colour compensated so that, on looking at a flat page, the reds appear to float above the page and, less usefully, the blues are sunk down.

It would also help if *Nature* were to advise authors (1) to code in spectral order and (2) to use blue and yellow for contrast, rather than red and green, where possible.

Mark Williamson

*Department of Biology,
University of York,
York, YO1 5DD, UK*

1. Kareiva, P. *Nature* **365**, 292–293 (1993).

Why the gloom?

SIR — There is a lot of gloom in the United Kingdom about the modest numbers of young people choosing to do physics, chemistry and so on. Why the gloom? There is no shortage of professional scientists and engineers. If there were, industry would pay them scarcity salaries. Research dependent on public funding suffers from too many claimants for a portion of the sizeable pie, not too few.

Where there is a real shortage is in able people educated through science (and therefore with sympathy for it) in responsible positions, working as bankers, administrators, politicians, managers and so on. But is our system at school and university geared to teach science in an educational manner? Most of the teaching appears to aim to produce professional scientists by working to overloaded and overspecialized syllabuses.

What is needed is to teach science as a human, social activity, to show how it evolved, to get across its spirit of adventure and its attitude, and to turn out graduates trained in communication skills and in teamwork. These are fundamental to all human activities, science and engineering included. No doubt the volume of science so taught in a given time would be less than now. Those who became professional researchers could make up for this deficiency later. At present many graduates starting on a research career first have to learn painfully to communicate effectively and to work with others.

I trust that more institutions will soon accept the need to teach science and engineering attractively and educationally, and make it known that they do so.

They would rapidly reap the benefit by attracting more and better students. Until this happens, they must be prepared to have smaller classes. Keeping up numbers by accepting students of low quality inflicts long-lasting damage on the standing and on the drawing power of science.

Hermann Bondi

*Churchill College,
Cambridge CB3 0DS, UK*

Not so new

SIR — Daedalus (*Nature* **365**, 300; 1993) thinks his “suckercar” is new but in fact it has been built, raced, and twice banned from further racing¹. The racing community preferred the more euphonious “hoover car”.

As in Daedalus’s suckercar, real hoover cars used the principle of a hovercraft in reverse. In the Chaparral 2J Can-Am racer, designed by Don Gates, two rear-facing fans were driven by an auxiliary 45 b.h.p. snowmobile engine. They sucked air from a chamber below the car which was sealed from the outside by skirts made of Lexan, a thermoplastic material. This held the car firmly to the ground allowing it to corner much faster than conventional cars. In this car Jackie Stewart made the fastest lap at Watkins Glen in July 1970 and Vic Elford was fastest in practice at several other Can-Am races. Unfortunately the car failed to finish in any race before the system was banned by the Fédération Internationale de l’Automobile (FIA) in December 1970 because it violated rules limiting the use of moving parts in aerofoil systems.

Another hoover car, designed by Gordon Murray, appeared on the Grand Prix circuit in 1978 and was more successful. The Brabham team’s BT46, based on the same principle as the 2J but with only one fan, was driven to an easy victory by Nicki Lauda in the Swedish Grand Prix and the result was allowed to stand although this system was also banned by the FIA.

Neither of the hoover cars was tested on vertical or inverted surfaces, as these are not usually a problem on Can-Am or Grand Prix racing circuits. The devices certainly improved cornering speed dramatically, and if the system had come into general use the drivers would probably have needed g-suits to protect them from the centrifugal forces.

Peter J. Bryant

*Developmental Biology Center,
University of California, Irvine,
Irvine, California 92717, USA*

Richard G. Knight

*General Refrigeration Limited,
Old Woods Trading Estate,
Torquay, Devon, UK*

1. Tilling, R. I. & Lipman, P. W. *Nature* **364**, 277–280 (1993).

2. Prata, A. J., Barton, I. J., Johnson, R. W., Kamo, K. & Kingwell, J. *Nature* **354**, 25 (1991).

3. Notes and procedures for Volcanic Ash Observations and Encounters ICAO Asia/Pacific (1993).

1. Boddy, W. & Laban, B. *The History of Motor Racing* (W. H. Smith, London, 1988).

Health research feels the chill

Barbara J. Culliton

Reforms in the US system for financing health care have grave implications for academic medicine and for basic scientific research.

US PRESIDENT Bill Clinton has made health-care reform (more specifically, reform of the health insurance industry) the centrepiece of the domestic agenda for his still new administration. Armed with the intelligence and charisma of his wife, Hillary Rodham Clinton, to whom he entrusted the task of devising a new plan for health care, the couple has come up with a beguiling proposal to extend insurance coverage and increase medical benefits for everyone in the country (including 37 million people who have no insurance now) without introducing new taxes.

How can they achieve this virtually cost-free miracle? By reducing waste and abuse in the current system, reducing unnecessary medical services, and forcing everyone in the country into a vast system of "managed care" medical collectives that, region by region, would negotiate favourable premium rates with insurance companies that want to stay in business. The Clintons are making the US public an offer that sounds too good to be true — largely because it is.

As they hit the trail of talk shows and popular appearances to sell their programme with rhetoric reminiscent of a political campaign, the Clintons give the impression that their reform plan will produce only winners. But this may not be the case. An army of lobbyists, journalists and others have scoured the legislative language of the 1,300-page draft reform bill on behalf of the elderly, small business owners, and middle-class tax payers who are likely to have to pay more than they do now in the Clinton land of health care. But, ironically and least expected, the most vulnerable group in this reform juggernaut may be the major medical schools (academic health centres) — the large, famous research institutions that have, for the past half-century, created and nurtured the vast US medicine and basic biomedical research enterprise.

Medical lobby

Leading medical researchers are now doing everything they can to place themselves in the centre of the debate that is likely to last at least a year and end in substantial changes to the Clintons' bill. Before the draft legislative proposal hit the streets, a handful of medical leaders, unofficially led by Michael Johns, dean of the medical school at Johns Hopkins, managed to hold a conference with Mrs

Clinton and her chief adviser, Ira Magaziner. The deans' influence is seen here and there in the 1,300-page draft bill, but this is just the beginning.

A foretaste of the debate to come was evident at a meeting last month of the Association of Academic Health Centres (AHC) in San Diego. This month, the Association of American Medical Colleges (AAMC) made health-care reform and its effect on education and research the focus of its annual meeting in Washington, DC. Mrs Clinton, who is taking to the academic hustings to promote the president's reform bill (she has spoken twice at Johns Hopkins, for example) gave the keynote address at the AAMC conference, where in a hard-hitting campaign-style speech she tried (not entirely successfully) to assure her audience of more than 1,000 deans and other medical leaders that the bill protects their institutions.

The promise that academic medical centres will get the financial support and subsidies they will need is by no means certain. The United States is entering an era in which low cost will be the be-all and end-all of health care. Research, high-technology medicine and innovation may be viewed as nothing more than costly accoutrements to a system that is already rhapsodically called "the best in the world" by its supporters.

There is a very real possibility that research will suffer unless the administration and Congress can come to understand the reasons why the support of academic medicine is in everyone's interest. For instance, a system that significantly reduces fees to physicians or hospitals, or that puts academic medical centres at a serious competitive disadvantage with respect to lower-cost community hospitals, will have the indirect effect of reducing the income that medical centres use to pay for education and research not fully covered by other funding, such as tuition, fellowships or NIH grants.

The language of health-care reform centres around the cost of insurance premiums, caps on doctors' fees, and reduced spending on Medicare and Medicaid — the two huge social programmes from the 1960s that pay (at reduced rates) the medical bills of the elderly and the poor. What has this kind of reform got to do with places like Harvard, Yale, Hopkins, the medical centre at the University of California at San Francisco (UCSF) and

other leading centres of biomedical research? Don't they depend upon federal grant-giving institutions like the NIH, which has an annual budget of \$11 billion, for their bread and butter? And is it not true that private philanthropies like the Howard Hughes Medical Institute (with \$7 billion in assets) can be relied on to enhance the contributions from the federal purse? What has a cap on Medicare reimbursements got to do with research in the clinical and fundamental biological sciences?

Symbiosis

The answer, alas, is plenty. Because of a prescient (or perverse) act of twentieth century history, the United States has evolved its own brand of medical education and science, centred around the research university with a medical school and one or more "teaching hospitals" at which young physicians and PhD candidates often work side-by-side. That point was made by NIH's next director, Harold Varmus, during his confirmation hearings before the US Senate on 3 November. Noting that a stint at NIH changed him from a would-be psychiatrist to a bench scientist, Varmus explained that he then spent most of his career as a professor in a basic science department of the medical school at UCSF, thereby underscoring the close link between medicine and basic research in the academic structure.

The institution that pioneers heart transplantation is also home to the discovery of recombinant DNA technology. In the United States, basic research is done in the medical centres. And because of the way in which these institutions manage their financial resources, basic science departments can be as dependent on the overall health of the hospital as are the more clinical departments, such as medicine, paediatrics or surgery. For example, most clinical departments have an income-generating "faculty practice plan" that varies from place to place in the details of its operation but has one essential element that is important to research: the dean's tax.

Although most of the income generated by physicians from patient care goes back to the doctors, a not insignificant tax (ranging from 5 to 15%) is taken off the top by the dean, who uses the money for various academic purposes. For instance, if a research project in molecular biology

is temporarily between NIH grants, the dean can take money from his tax pool to support the basic scientist through lean times. Therefore, it follows that if physicians' income is capped, the dean's tax will be smaller and the institution as a whole less flexible in discretionary funds.

It is a Byzantine system that few people (including researchers themselves) have really needed to understand until now. Briefly stated, the academic health centre is an organism with a vast array of hidden costs in which charges collected for one purpose are used to pay for something else. The trade jargon is 'cross-subsidization'.

Web of costs

Although cross-subsidies have been used to pay for what can be called hidden costs (largely because few people understand them), they are by no means secret or illicit costs. Accountants, federal legislators and insurance companies all agreed to play the game in which money was transferred from here to there to ensure the financial health of health centres where education, often-expensive patient care, and basic research were going on simultaneously.

The simplest example is the patient's hospital bill. By tacit agreement, insurance companies willingly paid an academic health centre as much as 130% of the actual cost of a routine procedure. Thus a gall bladder operation that might cost \$2,000 at a community hospital with no overhead costs for education or research would cost \$2,600 at an academic centre. The rationale is that academic medical centres are a national resource, and the extra bills they generate are equivalent to a tax of sorts to guarantee that the institution will be at the forefront of modern medicine for those who need it — heart and other organ transplantation, microsurgery, gene therapy and advanced cancer care, for example. The added cost might also go toward the cost of "uncompensated care" that major medical centres frequently offer to the poor and uninsured people living in their inner-city neighbourhoods.

But in the 1980s, as the cost of health insurance and care rose steadily (it is now 14 per cent of the gross national product), large employers took the lead in seeking ways to reduce costs, arguing that the sums they paid in insurance premiums for their employees put them at a competitive disadvantage internationally. And so what is now called 'managed care' began to grow. Originally known as Health Maintenance Organizations, health-care groups were formed to provide a certain predetermined amount of care to every member of the group at a fixed (reduced) fee. This immediately put more expensive academic health centres at a competitive disadvantage. Congress, too, got into the

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Emphasis on health — Hillary Rodham Clinton testifying before Congress during hearings on health-care reform in September. Since then, accompanied by the popular former Surgeon General C. Everett Koop, she has been selling the proposals to doctors at a series of forums.

act. Since the inception of Medicare in the 1960s, academic medical centres have received a set education cap for every Medicare patient treated — another of the hidden, but not secret, taxes in the system.

But since the 1980s, Congress has been instituting in law a series of gradual reductions to the Medicare subsidy. It did not take academic medical leaders long to figure out that the system of cross-subsidization would have to be replaced by a more coherent, direct system of support for the training and research distinguishing the big medical centres from community hospitals that very often offer equally good patient care, especially in routine cases. For instance, in 1985 the Commonwealth Fund supported a study headed by Robert M. Heyssel (then president of the Johns Hopkins Hospital) that in essence asked the government fully to acknowledge its commitment to academic medical centres by replacing cross-subsidization with direct funding. That is in part what the Clinton bill would do. Speaking at the AHC meeting last month, Igal Paroo, head of Hahnemann University in Philadelphia (an AHC institution), put things in perspective when he said "The Clinton plan is just the tail-wind accelerating changes that are already taking place, driven by market forces." The Clintons' reform package (however it turns out in the end) will make change a reality.

Certain provisions of the draft legislation deserve special comment and raise important questions.

Innovation

There is no question that innovation is

expensive. First, it can increase patient demand. Although advances in ophthalmology, for example, have made day or out-patient eye surgery routine, the savings incurred from innovative surgical techniques are offset by an increase in patients having surgery. Second, innovation is often expensive to accomplish and to deliver. Hopkins' surgeons recently reported significant improvement in an artificial larynx for cancer patients and others who have lost the ability to speak. Will cost-conscious business directors of managed-competition health plans want to pay for the improved larynx, or will they say the older procedures for helping patients are good enough? Will the system thereby discourage the very kind of innovation that has made US science so dominant? The Clintons' plan does not adequately deal with the threat to innovation.

Direct payment for education

The reform bill does make a stab at the issue of support for graduate medical education — the training of medical school graduates in their residency years, which usually leads to specialization. Wisely, the bill takes the burden from Medicare alone and makes provisions for all-payers (government and private insurers) to contribute to a direct graduate medical education 'account' to pay salary and benefits for about 15,000 men and women presently in training. In 1993 that bill comes to about \$2 billion.

The draft bill also includes money to pay for the indirect costs that the medical centres' teaching hospitals incur because they have a high rate of non-payment and because they treat patients who are often

sicker and far more in need of high-technology medicine than those in community hospitals. This year these costs are calculated at about \$3.8 billion, from a total of \$5.8 billion for education-related costs. The Clintons' plan will phase in direct cost reimbursement over a period of 5 years, calling for expenditures of \$5.8 billion for the education account by 1998 and payments of \$3.2 billion for the indirect costs by the same year. This is not nearly enough.

Primary care vs specialization

The Clintons' plan reflects a belief frequently reiterated by academic medical groups during the past decade — that a health insurance programme promising universal access to care requires a significant increase in the number of primary-care physicians and some reduction in the numbers of specialists that are now dominant.

The implications for change in this area are staggering. First, for 50 years the US medical system has been guided by a deliberate policy encouraging specialization. Linking medical schools so closely to research universities (rather than establishing them as free-standing trade schools) has fostered an ethic that places at the top of the pyramid the physician who is equally adept at patient care, teaching and research — the latter leading naturally to specialization. Second, by design (right or wrong), specialists are paid more than generalists, which places an additional premium on specialty medicine — particularly for students who graduate from medical school thousands of dollars in tuition debt.

Changing this ethos will require a sea-change in the way doctors are educated and among the mentors they are taught to emulate. Some of the leading medical schools are already trying to establish the general internist as an equal member of the country's large and powerful departments of medicine, from whose ranks have come many renowned basic scientists. Giving generalists equal status will be a tough job. Figuring out how to change medical curricula to make a doctor both good at the care of routine diseases and knowledgeable about the new 'molecular medicine' that will soon come into its own is another challenge; genetics is an obvious example. The number of known disease genes for which tests are available is growing quite literally week by week. But medical genetics is not now a major part of the general curriculum, and there are only 600 recognized specialists in this field in the United States. Furthermore, there is no way of knowing where genetics, as the first wave of molecular medicine, will stand in the cost-cutting scheme of managed competition. Nor, for that matter, do we know that the plan to increase the proportion of generalists

among medical graduates will actually improve the overall health of the public and reduce costs, which is the whole purpose behind the Clintons' idea of the generalist movement.

The élite schools

Americans like to think of themselves as an egalitarian lot, so it is not surprising that the original version of the reform plan would have required that 55% of medical school graduates (school-by-school, or region-by-region) elect a career in primary care, which includes general internal medicine, paediatrics, and obstetrics and gynaecology. (The list may grow longer as other groups who now call themselves specialists argue for a primary-care designation: ophthalmology, for instance.) But, thanks to effective persuasion by academic leaders, the draft bill now requires only that 55% of the country's medical graduates as a whole enter primary care, lured into the field by enticements such as tuition loan forgiveness not available to physicians taking the path to specialization and research.

Only about one-third of the academic medical centres are research-intensive institutions, which receive the lion's share of NIH's grant money. It would be a great waste of research talent to force all of the United States' medical centres into the same mould. Thus, if the current legislative language holds, the best academic institutions might graduate 20% of their classes in general medicine, while the proportion going into primary-care fields at less research-intensive centres might well be higher.

Nonetheless, it is vital to the success of the primary-care movement that the élite research institutions do their part to take the lead in developing innovative ways of teaching primary care in a world now intellectually dominated by 'molecular medicine'. It is equally important that the élite institutions stick to their current commitments to make primary care as important to young physicians as the more traditional route of research and specialization.

Reducing faculty numbers

If academic health-care centres deserve special consideration under health reform (and they do), these centres must also recognize that government policy encouraging expansion has allowed them to become bloated — expensive beyond what is justifiable in terms of special patient care or of research. During the past decade, the numbers of medical school faculty overall have climbed from 60,000 to 70,000. Some undefined but generally agreed to be large proportion of the extra 10,000 are physicians hired by clinical departments to generate more patients and, therefore, more income. Many of these physicians neither teach

nor do research and, as things emerge in the new system, they will not be considered a legitimate part of the institution's academic overhead.

All across the country, deans are trying to figure out how to get rid of this excess faculty, a problem compounded by the tenure system which may also change in the next decade. As things stand now, however, Robert G. Petersdorf, president of the AAMC, got it right when he said, "The only good thing you can say about tenure is that where there's death there's hope." Overall, the message of health-care reform to academic medical centres is that, despite some special financial considerations, they will have to learn to make the most of the idea that less is more.

Static budget for NIH

"Downsizing", the business community's euphemism for a reduction in force, may also occur in science because of an expected standstill in the budget for the NIH, now at an all-time high of \$11 billion. (This is, in fact, only \$500 million more than requested by the Bush Administration.) A legally binding cap on discretionary federal spending for the next 5 years puts NIH in direct competition as never before with such other worthy social goods as Head Start (an early education programme for the poor), homes for the homeless, and educational retraining of defence workers displaced by the end of the Cold War. Although NIH's staunchest supporters press Congress for a doubling of the budget, this is not likely to come about.

Where next?

Health-care reform is certain to dominate the political agenda well into 1994, if not beyond. In addition to the Clinton bill which has received so much attention, alternative bills are also before the Congress. Further, the president has made it clear that he is willing to negotiate on virtually all of the specifics, except universal coverage.

The president's proposal makes assumptions about the cost of health reform that do not add up. Despite the administration's insistence that universal coverage and a basic benefits package can be paid for largely out of cost-savings (from making the system more efficient and competitive) and a modest 'sin' tax on cigarettes, it is highly likely that Mr Clinton has greatly underestimated the cost of health care reform. Therefore, one can expect taxes, cost projections, and the effects on the economy to dominate the forthcoming debate. It will be easy for academic health centres to get lost in the shuffle, but that must not be allowed to happen. □

Barbara J. Culliton is Deputy Editor of *Nature*.

The changing shape of structure

Participants at *Nature's* conference on structural biology (11–12 November) found that mere description of structures has given way to an exciting exploration of the relationship between form and function.

Boston. The term "structural biology" was coined about 25 years ago. For much of the intervening time, however, progress has been slow, so much so that, when the conference began last week, John Maddox expressed concern that the subject had yet to progress beyond the merely descriptive. He need not have worried. Confidently announcing a new central dogma for the Age of Structural Biology ("sequence implies structure implies function"), Greg Petsko (Brandeis University) introduced the first session by noting that progress in the field was now so fast that the important thing was no longer who had described what first, but what they had done with it.

The acceleration has indeed been dramatic — this year, new structures have been appearing at the rate of almost one a day. As Wayne Hendrickson (Columbia University) explained, the change stems from a revolution in the available instrumentation. Synchrotron radiation sources and area detectors have drastically reduced the time taken to gather the data for a new structure, and are allowing techniques such as multi-wavelength anomalous dispersion (MAD) to become routine. Coupled with vastly enhanced computer power and the large quantities of pure protein that can now be made by recombinant DNA techniques (sometimes containing atomic labels such as seleno-methionine), this has led to an exponential growth in structure numbers.

As the flood of new information continues, it is increasingly common to find that two seemingly unrelated proteins (such as interleukin 1 β and soybean trypsin inhibitor) turn out to have the same fold. Is there any way in which such pairs can be recognized? Janet Thornton (University College, London) described a remarkably successful method for "threading" primary sequences on to folds known from other proteins and evaluating how well they fit. Of course, at least one protein with the relevant fold must have been analysed already. But there may be only about 1,000 folds to discover, as Cyrus Chothia (University of Cambridge) predicted, and as the number of known folds (currently about 135) increases towards this limit, the technique will be increasingly useful.

Other talks concentrated on even more intractable problems. Some of the hardest are concerned with the motions and actions of proteins, which frequently occur in milliseconds, rather than the minutes or hours needed to obtain a complete set of X-ray diffraction data. Sometimes intermediate forms can be trapped as Thomas Creighton

(EMBL, Heidelberg) has done to study the folding of bovine pancreatic trypsin inhibitor. On other occasions, however, new approaches are necessary. Petsko's group has been using the Laue diffraction technique, in which data collection is accelerated by using broad bandwidth X-rays on a stationary crystal, to examine short-lived intermediates such as the GTP-bound form of the oncoprotein Ras.

In other cases again, even higher time resolution is possible. Structure determination by NMR is now a much used alternative to X-ray diffraction. Kurt Wüthrich (ETH, Zürich), however, showed that it can also be used to look at the water molecules trapped between a homeobox DNA-binding domain and the major groove of the DNA, demonstrating that they are freely exchangeable on millisecond timescales.

Studying the structures of proteins that are normally inserted in membranes entails further problems, as Johann Deisenhofer (University of Texas Southwestern Medical Center) noted, not least those of obtaining an adequate supply of pure material and persuading it to form usable crystals. It is therefore hardly surprising that only a few such structures (notably that of the photoreaction centre of *Rhodospseudomonas viridis*) have ever been determined. Does this mean that any hope of understanding the motions such proteins undergo must be abandoned? Not at all, said Nigel Unwin (MRC Laboratory of Molecular Biology, Cambridge). By using electron rather than X-ray diffraction, and by spraying acetylcholine onto two-dimensional crystals of its receptor just milliseconds before freezing them, his group is well on the way to understanding how the binding of this neurotransmitter opens ion channels in the postsynaptic membrane.

Sometimes, the protein itself helpfully remains in an intermediate state for long enough to be lopped off its membrane anchor and studied. Don Wiley (Harvard University) described a striking example from influenza virus, whose haemagglutinin undergoes an irreversible structural rearrangement on exposure to the acidic environment of lysosomes. The acid-induced structure is produced by several concerted helix-stand transitions, and extends an α -helix in the extracellular form of the protein by more than 100 Å so as to expose the highly hydrophobic N-terminal peptide and induce membrane fusion. A similar mechanism may be used by other enveloped viruses, suggested Stephen Harrison (Harvard Univer-

sity), but at least for flaviviruses the detailed mechanism appears to be very different.

Can such an intimate knowledge of the viral life-cycle be used to interfere with it? Mark von Itzstein (Monash University, Melbourne) showed that it can. By a careful exploration of the sialic acid binding site on the influenza virus sialidase (also known as neuraminidase), he and his collaborators have been able to synthesize sialic acid derivatives that bind tightly and specifically to the influenza enzyme alone and show great promise as antiviral agents. Other groups have been developing methods by which similar inhibitors can be designed for any enzyme. By cross-linking crystals of elastase and transferring them to a variety of organic solvents, Dagmar Ringe (Brandeis University) has been able to map any molecules of each solvent that are tightly bound to the protein surface, building up a picture of the places available for interaction with any possible inhibitor.

But in the end, it was the sheer power of the insights into biological problems that structural methods are now producing that was most impressive. Our understanding of how muscles contract (Ken Holmes, Max Planck Institute, Heidelberg), how transcription factors bind DNA (Aaron Klug, MRC Laboratory of Molecular Biology, Cambridge), how calcium levels control enzyme activity (Florante Quiocho, Baylor College of Medicine, Texas), and how tyrosine phosphorylation transmits intracellular signals (John Kuriyan, Rockefeller University, New York) now rests upon the structure of the proteins and substrates involved, and in the past twelve months, our views of peptide presentation by the immune system and the eukaryotic transcription initiation have been transformed by the structures for peptide-bound MHC class II (Don Wiley, Harvard University) and DNA-bound TATA-box binding protein (Stephen Burley, Rockefeller University, New York) respectively.

Who would have imagined without seeing the structures involved that, despite the different roles, subunit compositions, and peptide binding capacities of MHC class I and II, their peptide binding domains would have virtually identical structures, or that a basic component of the transcription apparatus could so drastically affect the structure of its binding site? As the remaining 900 folds are discovered and their functions are explored, such insights are sure to become increasingly frequent, ensuring that the effects of the structural explosion are felt for a long time to come.

Nicholas Short

Small and perfectly formed

Peter Little

WHAT has the pufferfish, *Fugu rubripes*, got to offer the modern biologist? Sydney Brenner and colleagues let on elsewhere in this issue (*Nature* **366**, 265–268; 1993), where they propose that it should be added to the select list of surrogate organisms used in research. Much of our understanding of basic human biology stems from work with such organisms, which are used to overcome the practical, ethical and social problems of experimenting on human beings. The main surrogates are few in number, examples being the bacterium *Escherichia coli*, yeasts, nematode worms, fruitflies, clawed toads, zebra fish and house mice.

Fugu is widely, if expensively, consumed in Japan but it cannot be bred in captivity nor is it found in European waters; it has no known mutants and in all respects seems an unlikely choice to put on to the list of surrogate species. What prompts Brenner *et al.* to argue for its inclusion is that this fish has the smallest genome of any vertebrate, and with current molecular genetic technology this may have important implications. The problem with mammals is that they have large genomes full of nongenic DNA, such as repeats, pseudogenes and other noncoding sequences, which makes genome analysis difficult and labour intensive. Many of the technical problems might be eliminated by studying *Fugu* DNA because its genome is about seven times smaller than that of the human or mouse.

Brenner *et al.* estimate genome size by two methods. Most simply, they analyse the frequency with which unique sequences can be detected, by hybridization, in unamplified phage libraries containing *Fugu* genomic DNA. A given DNA fragment was found on average once in 24,625 phages which, with an insert size of 16.5 kilobases, gives a total genome size of 404 megabases.

The second method is more complex and involves sequence analysis of random genomic DNA fragments. The authors reasoned that the frequency with which genes could be detected in these sequences (using computer analysis of protein databases), compared to the frequency with which genes could be detected in human DNA using similar database analysis, could be used to calculate the relative size of the *Fugu* genome. They sequenced 127,831 base pairs of *Fugu* DNA and

1,011 of them (0.791 per cent) coded for protein. In contrast, computer analysis of existing protein-coding sequences suggests that random sequences of human (or mouse) DNA would contain 0.103 per cent coding sequences — a 7.68 times enrichment in *Fugu*. To a good first approximation humans and *Fugu* have

is unexpanded and never acquired large quantities of 'junk' DNA. It may therefore more faithfully reflect the original vertebrate genome, uncluttered by sequences that have accumulated in the evolutionary line leading to contemporary mammals". Testing this hypothesis should be possible by systematic genomic taxonomy.

If one had to describe the *Fugu* genome in a few words, one would say that it was like the human genome but with less junk. This is the principle on which Brenner *et al.* base their proposal that *Fugu* could become a member of the group of surrogates for mammalian species.

Their argument is not yet completely compelling, but experiments now being carried out will determine its limitations. In brief, from sequence analysis it seems that *Fugu* genes are proportionately small because they have small introns (80-base-pair median size). So gene analysis in *Fugu* will be much simpler than in a mammal because the introns are so much smaller. Similarly, finding new genes will generally be easier because, quite irrespective of the methods employed,

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The pufferfish, *Fugu rubripes rubripes* and its genome compared to that of the human. Human chromosomes 1 to 12 are displayed in the upper row, 13 to X and Y below. In size, the pufferfish genome would occupy the shaded box covering human chromosome 1 and about half of 2.

the same number of genes, so it must follow that the *Fugu* genome is 7.68 times smaller than that of humans, giving a genome size of 390 megabases. The sequence analysis also shows that 7.6 per cent of the genome is repeated. Most important, the authors could see no sign of the interspersed highly repetitive sequences that are characteristic of human and mouse DNA, meaning that in the *Fugu* genome most of the repetitive DNA is probably clustered.

An interesting side issue is why *Fugu* has such a small genome. Did the species lose all the intergenic 'junk' or did it never have such DNA in the first place? Brenner's opinion on this is that "Rather than being contracted, the *Fugu* genome

one will have to put some 7.68 times less DNA into any analytical protocol. Preliminary analyses in Brenner's laboratory support both of these views.

Much more exciting is the possibility that gene order in *Fugu* is conserved compared to that in humans; at present there is limited information about long-range gene order, but certainly conservation of short-range linkage has been observed. If conservation extends to clusters of five or six genes then the *Fugu* surrogate could have a dramatic effect upon positional cloning strategies. Imagine having markers flanking a region in the human known to contain a disease locus: if linkage can be demonstrated in *Fugu*, then the amount of DNA to be

sifted through to find candidate gene(s) becomes about 7–8 times less. In practical terms, this would reduce a walk using yeast artificial chromosomes (YACs) to a cosmid-based one, which, as a technical restructuring of a research project, would be a dramatic change and gratefully accepted by anyone familiar with YAC clones and the frustrations they cause.

A further experimental step has to be overcome: as yet, no one knows if *Fugu* genes work in other organisms. This experiment (currently under way) is crucial because analysis of transgenomes is the

potential key to the genetics of *Fugu*. As Brenner *et al.* point out in the closing sentences of the paper, transgenesis is the “means to assaying the value of the sequence and adds the necessary functional dimension to genome research”. Their work is an elegant expansion of the rationale for genome research and the *Fugu* proposal deserves attention from any thoughtful biologist. □

Peter Little is in the Department of Biochemistry, Imperial College, Prince Consort Road, London SW7 2AZ, UK.

GEOMAGNETISM

A swiftly changing field

Robert Coe

THE average intensity of the Earth's magnetic field last peaked, roughly speaking, at the birth of Jesus, when it was 40 per cent stronger than today. Around the globe, the instantaneous peak arrived as much as 800 years earlier or later, thereby encompassing the births of Buddha and Mohammed as well. Since then it has decreased systematically, and continues to plummet today at 7 per cent per century, inspiring speculation that the field is heading towards zero and a polarity reversal (although not in time for the new millennium). On a much longer timescale, evidence for a 100-million-year interval of unusually low intensity in the Mesozoic has fuelled speculation about connections between the geodynamo, episodic mantle convection, and rates of sea-floor spreading, hotspot volcanism and true polar wander^{1,2}. But data have been sparse. Now on pages 234 and 238 of this issue, two innovative palaeointensity studies are reported, one using submarine basaltic glass³ and the other using a great length of deep-sea sediment core⁴.

Why, given its diagnostic potential, have there been so few studies of palaeomagnetic field intensity? The reason is that the intensity of magnetization of rocks depends on many more factors than does its direction. To obtain the actual intensity of the ancient field that magnetized a given sample, one must duplicate in the laboratory the essential elements of the natural process by which the magnetization was acquired. This has been feasible only for lava flows and archaeological objects such as pottery, which acquire a thermoremanent magnetization at the time they form by cooling quickly from above the Curie point of their magnetic minerals to ambient temperature in the Earth's field. Because it is a linear function of intensity in weak fields, the ratio of natural thermoremanence to that acquired by cooling the same sample in the laboratory should, ideally,

give the ratio of ancient to laboratory field.

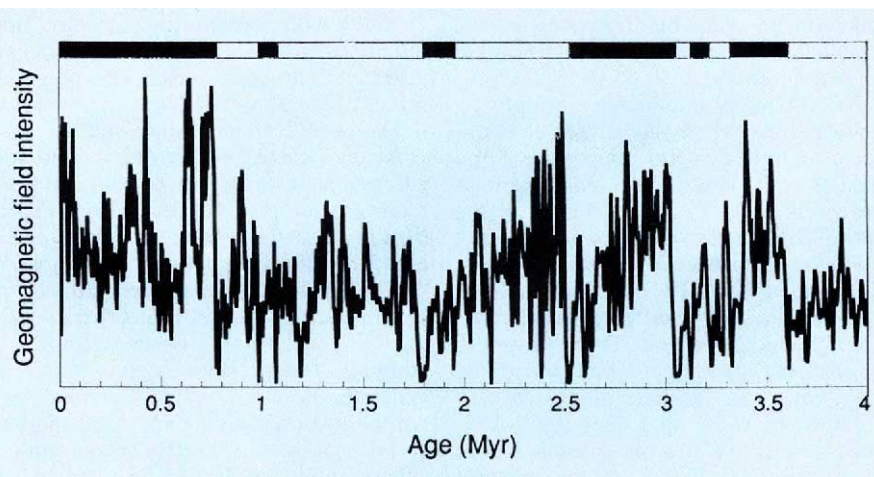
Sadly, the world is often far from ideal. The primary thermoremanence of lava flows may be obscured by later magnetizations of diverse origin, a problem that generally becomes more serious with increasing geological age. Equally serious is alteration of the primary magnetic minerals during heating in the laboratory, a problem that affects young and old rocks alike.

Pick and Tauxe³ have discovered that small fragments of fresh basaltic glass which behave nearly ideally in palaeointensity experiments can sometimes be obtained from the quenched rind of submarine basalt flows. Their finding flies in the face of conventional wisdom that sea-floor basalt is unsuitable for determining field intensity because of the rapid hydrothermal alteration it undergoes shortly after formation at spreading ridges. To do the experiments, they developed new handling techniques, overcoming problems that arise from the tiny

size and significant magnetic viscosity by enclosing the samples in cold-pressed salt and by measuring them repeatedly in zero field in a very sensitive cryogenic magnetometer. The magnetic carrier, they find, is submicroscopic, single-domain, titanium-poor magnetite, somewhat surprising in view of previous observations of titanium-rich magnetite in fresh pillow basalts. Might the magnetic mineral in the submarine glass be secondary, and the magnetization acquired after cooling when the new mineral formed? Pick and Tauxe effectively dispose of this alarming thought by showing that their results for samples erupted in 1990 on the East Pacific Rise successfully reproduce the known field intensity there.

Turning their attention to the low-field interval in the Mesozoic, they find low intensities, 25 to 45 per cent of what would be produced by today's dipole field, in basaltic glass samples 85 to 125 million years old, from four cores drilled in the Atlantic, Caribbean and Pacific oceans. Three of these cores bracket the onset of the Cretaceous normal superchron, a 40-million-year interval of no magnetic reversals, and the fourth falls just before its end. If the low Mesozoic field extended throughout this superchron, as Pick and Tauxe contend from their results, this demolishes a train of theoretical reasoning linking the extraordinary stability of the field to high dipole intensity^{5,6}. This conclusion, however, excludes other higher palaeointensity results (average 65 per cent of the present dipole field) from Indian and Israeli lava flows that lie between theirs in age, but which admittedly yielded intensity estimates of lower quality. It also indicates a very rapid shift at the close of the superchron to the higher-intensity regime that continues today (see their Fig. 3).

In contrast to the isolated readings provided by lava flows, sedimentary rocks offer a much more continuous magnetic



'Sawtooth' variations in the intensity of the Earth's magnetic field over the past four million years, derived by Valet and Meynadier (see page 234 of this issue). Dark/light horizontal bars indicate normal/reverse polarity.

record. But because we don't know how to duplicate the process by which they acquired their natural magnetization, only relative intensity changes can be determined, by normalizing the natural magnetization with a suitable factor to remove the influence of all variables except field intensity. Chances of success are improved by using only sediments with simple mineralogy, laid down under fairly uniform conditions, and by choosing the normalizing factor wisely⁷. Valet and Meynadier³ have applied this technique on an unprecedented scale, obtaining a detailed record of relative palaeointensity that spans the past 4 Myr and 11 reversal boundaries, by analysing in detail 80 metres of sedimentary core from the equatorial Pacific Ocean. Technological advances, allowing treatment and measurement of continuous segments of core with a horizontal pass-through cryogenic magnetometer, let them complete many tens of thousands of measurements in a practical length of time. Three different normalizing factors yielded similar results.

Their record reveals an intriguing 'sawtooth' pattern (see figure on previous page) that has not been recognized previously: namely, a gradual decrease in average intensity during times of constant polarity punctuated by a sharp recovery just after polarity reversals. In addition, it seems that the higher the recovery of the field intensity following a reversal, the longer is the interval of stable polarity until the next reversal. Valet and Meynadier's interpretation is that the usual state of affairs in the core is for the convective regime driving the geodynamo to decay, and the dipole field to lose strength, until dynamo action is rejuvenated by a polarity reversal. Low field intensity would promote polarity reversal, much as Allan Cox envisaged more than 20 years ago⁸. Indeed, studies of sequences of lava flows spanning polarity boundaries have detected the same asymmetry in intensity before and after the reversal, although with a smaller jump in relative intensity⁹.

As with many simplifying conceptions, there are pieces of evidence that do not as yet seem to fit this enticing picture. For a start, why, if the average intensity were as low as Pick and Tauxe concluded for the entire Cretaceous 'quiet period', didn't the field reverse even once? Perhaps it simply means that the convective regime in the core was quite different then, as has often been suggested. Then, too, the average dipole moment that Valet and Meynadier infer for the past 4 Myr — obtained by calibrating their relative intensity record against the volcanic record for the past 140 kyr — is only half that found by averaging results from lava flows erupted during the past 5 Myr. Inhomogeneous distribution of the volcanic

data in time probably explains part of the discrepancy, but not all of it. And the present polarity interval does not show the expected gradual decrease in intensity in the sedimentary record. Might the sawtooth pattern be caused by partial realignment of magnetic grains induced by the subsequent field of opposite polarity? The realignment would be expected to decrease progressively below the interface between sediment and water, as compaction and cementation increasingly hamper grain rotation. Reassuringly, though, the depth-range over which grain rotation would have to occur to explain the record is incredibly large, 10 to 100 times greater than current estimates.

As the authors point out, the present polarity interval is different from the others in several regards; perhaps most relevant is the abundance of geomagnetic excursions (see Valet and Meynadier's

Fig. 3), where the field changes direction and often drops suddenly in intensity, but does not stably reverse before returning to the original polarity.

Such unanswered questions about geomagnetic intensity variations and their relation to the dynamics of the Earth's deep interior are no cause for pessimism. My guess is that many of them will be answered before the next millennium. □

Robert Coe is in the Department of Earth Sciences, University of California, Santa Cruz, California 95064, USA.

1. Prévot, M. *et al.* *Earth planet. Sci. Lett.* **97**, 129–139 (1990).
2. Merrill, R. T. *Nature* **345**, 575–576 (1990).
3. Pick, T. & Tauxe, L. *Nature* **366**, 238–242 (1993).
4. Valet, J.-P. & Meynadier, L. *Nature* **366**, 234–238 (1993).
5. Pal, P. & Roberts, P. *Nature* **331**, 702–705 (1988).
6. Larson, R. L. & Olson, P. *Earth planet. Sci. Lett.* **107**, 437–447 (1991).
7. Tauxe, L. *Rev. Geophys.* **31**, 319–354 (1993).
8. Cox, A. V. *Science* **163**, 237–245 (1969).
9. Prévot, M. *et al.* *J. geophys. Res.* **90**, 10417–10448 (1985).

BIOCATALYSIS

Charge of the light brigade

M. Grätzel

LIGHT-induced charge separation in colloidal semiconductors has attracted the attention of the scientific community for more than a decade¹ because of its potential uses in photocatalysis and optoelectronics. The challenge is to identify systems that avoid the energy-wasting electron-hole recombination reaction which is known to proceed in such small particles on a timescale of nanoseconds. Intercepting the back reaction is a prerequisite to obtain high quantum yields for photodriven redox processes in colloidal semiconductor systems. A new strategy used by Willner and colleagues² is to scavenge the conduction-band electrons produced by light excitation of semiconductor particles (colloidal titanium dioxide) with an electron acceptor immobilized close to the surface by covalent attachment to a polymer. The polymer serves at the same time as a support for a redox enzyme which reduces nitrate.

At the heart of the device is a microheterogeneous assembly of colloidal titanium dioxide particles, each surrounded by a redox-active polyethylenimine polymer. The role of the polymer is to stabilize the semiconductor particles, which in the absence of such a 'protective agent', would flocculate in aqueous solution. In addition, the polymer supports methyl viologen, the acceptor for the conduction-band electrons produced. The methyl viologen is bound directly to the amine groups of polymer backbone by an amide bond. Organized in this way, the system achieves a high local concentration of the electron acceptor at the semiconductor

surface, so interfacial charge transfer from the conduction band of the colloidal semiconductor to the acceptor is rapid. Light excitation of the assembly efficiently forms the singly reduced form of viologen, MV⁺, which Willner and colleagues detected optically by monitoring the build-up of its characteristic absorption peak around 600 nanometres. The high quantum yield for MV⁺ formation observed shows that the interfacial electron transfer can compete successfully with electron-hole recombination.

Willner *et al.* illustrate impressively the advantages of their system by performing for comparison photoreduction experiments with polyethylenimine-protected TiO₂ colloids, where the viologen was not attached to the polymer, instead being allowed to move freely in solution. In this case, the electron acceptor has to diffuse to the surface of the TiO₂ particles before it can scavenge the conduction-band electrons, and this greatly reduces the efficiency of light-induced charge separation. The rate of methyl viologen photoreduction was reduced to about a fiftieth, and the quantum yield of MV⁺ formation, measured by laser flash photolysis, to a mere fifth of that for the polymer-bound viologen. Avoiding the diffusional displacement of the viologen by surface attachment to the polyethylenimine seems to be the key to greatly enhanced interfacial electron transfer.

As a further illustration of the practical uses of this approach, the team describes a third type of experiment where the primary charge separation involving viologen

Why we still love Lucy

SPECIMENS of *Australopithecus afarensis* from 3.4-million-year-old deposits in the Middle Awash at Maka, Ethiopia, described by Tim White and colleagues on page 261 of this issue, should lay an acrimonious debate to rest. Fourteen years ago, Johanson and White surprised (and exasperated) many by basing their new taxon *A. afarensis* not on the Ethiopian skeleton nicknamed 'Lucy', but on more robust and stratigraphically more ancient specimens collected far to the south, at Laetoli in Tanzania, by Mary Leakey and her associates. The implication was that *A. afarensis* was a long-lived species with somewhat variable morphology and high degree of sexual dimorphism.

Critics pointed out that the Ethiopian specimens represented smaller individuals, in general, than those in the Tanzanian collection, begging the question of why males and females should have lived in different places. But the critics went further — the Tanzanian forms lived in fairly open country, and their bipedality is confirmed by the presence of fossil footprints. Lucy's neighbourhood, in contrast, was probably more thickly wooded, with the implication that Lucy might have been a tree-dwelling species.

The cynic will spot a hidden agenda here. The Leakey school stresses the antiquity of



Homo, and for others to have muscled in on the Laetoli fossils and named them *Australopithecus* was a source of contention. In

this light, it seems natural for Johanson and White's critics to stress the bipedality, larger size and open-country life of the Tanzanian forms, casting the (more recent) Ethiopian forms as arboreal and irrelevant offshoots of the human lineage.

The new finds from Maka could even the score (inasmuch as any score can be truly settled in a discipline where ego so often triumphs over evidence). The specimens are intermediate in age between Lucy and those from Laetoli, but display the full range of size seen in both collections. The large jaw in the accompanying picture (top) is MAK-VP-1/12 from Maka, and is the most complete mandible known from the species. The smaller one (bottom) is AL 288-1 from Lucy herself. So there is no doubt that *A. afarensis* is a single, variable species. An accompanying humerus (see Fig. 2a on page 264), although clearly belonging to the same species as Lucy, is large and robust, suggesting that it belonged to a male: the degree of sexual dimorphism is probably no greater than that seen in the extant gorilla (or the extinct *Lusengpithecus*). But the

humerus is also relatively short, which is not what one would expect in a tree-living creature. Henry Gee

Photo by Tim White

photoreduction is coupled to enzymatically catalysed reduction of nitrate to nitrite. The coupling is neatly achieved by covalent bonding of the enzyme nitrate reductase to the polyimine polymer. The latter serves to transport electrons between the TiO_2 particle and the enzyme. Electron flow occurs in cascade fashion: the conduction-band electrons reduce the viologen which in turn reduces the nitrate reductase. Two electrons are pooled by the catalytically active group of the enzyme before they are passed to the nitrate substrate to yield nitrite as a final product.

The concept is not new; electroactive polymers of similar type have been applied previously³ to ensure electronic charge transfer between macroscopic electrodes and enzymes. But the results of Willner and colleagues open the way to extend this type of redox 'wiring' to colloidal assemblies of semiconductors and enzymes.

Apart from the reduction of nitrate, photobiocatalytic semiconductor nanocrystals may be applicable to other multi-electron transfer reactions that require the

intervention of natural enzymes or artificial redox catalysts. An example is the reduction of carbon dioxide. The efficiency of these photocatalytic reactions is often poor, and using redox-wired semiconductor/enzyme assemblies may well lead to improvement of the quantum yields. There are strong hopes of generating fuels such as methanol or fertilizers such as ammonia in this way using sunlight.

Instead of dispersing the colloidal semiconductor particles in solution, one can deposit them as a layer on a conducting glass support. Such nanocrystalline films exhibit some extraordinary properties. For example, colloidal TiO_2 films whose surfaces were treated with a charge-transfer sensitizer are extremely efficient at photovoltaic conversion of solar energy⁴. Another useful feature is that the Fermi level within the particles can be potentiostatically controlled by incorporating the supported film as the working electrode in a conventional three-electrode electrochemical cell. This opens up the prospect of using assemblies of the

type reported by Willner and colleagues as catalytic enzyme electrodes. Multi-electron reduction processes could then be accomplished in the dark, the particle layer being used as a cathode. And indeed, studies in progress at University College Dublin (D. Fitzmaurice, personal communication), using molecular triads (which are composed of three redox-active molecules, with a gradient in redox potential from the first to the third) bound to the surface of nanocrystalline TiO_2 films, confirm that vectorial electron flow occurs and that efficient charge separation can be obtained with such systems. □

M. Grätzel is at the Institut de Chimie Physique, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.

1. Grätzel, M. *Heterogeneous Photochemical Electron Transfer* (CRC, Boca Raton, 1989).
2. Willner, I., Eichen, Y., Frank, A. J. & Fox, M. A. *J. phys. Chem.* **97**, 7264–7267 (1993).
3. Heller, A. *Accs chem. Res.* **23**, 128 (1990).
4. O'Regan, B. & Grätzel, M. *Nature* **353**, 257–260 (1991).

The writing is on the wall

S. F. Blinkhorn

JUST as the handwritten letter is about to succumb to competition from the word processor, there seems to be a growing trend for companies to ask potential recruits for applications completed in their own handwriting. This is nothing to do with a sudden upsurge in aesthetic sensitivity on the part of personnel managers. One need look no further for an explanation than the marketing activities of graphologists.

Graphologists claim to be able to discern character and suitability for employment from handwriting. The claim finds a ready audience in the higher reaches of business, in particular in financial institutions. In continental Europe — especially France — the influence of graphologists is reputedly pervasive, although credible statistics are hard to come by and organizations that are believed on good grounds to use their services will often refuse to acknowledge the fact.

These practitioners, who are organized in competing professional associations, present their findings as matters of fact not opinion and lay claim to diagnostic skill of a high order. A statement produced today by the British Psychological Society (BPS)* presents a contrary view. Drawing together findings from published empirical work, it presents an unremittingly negative picture of the value of graphology in personnel assessment. Copies of the statement have been sent to the Secretary of State for Employment, the House of Commons Select Committee on Employment, the Commission for Racial Equality, the Equal Opportunities Commission and the Institute of Personnel Management, in the hope of opening up the issue of whether the use of graphology constitutes unfair employment practice.

It is not for want of trying that no persuasive evidence has been found. Researchers have tried to find associations between the judgements of graphologists and scores on a variety of psychological tests. They have looked for relationships with the criteria used in a full-scale assessment centre being run live as part of a selection procedure, and with supervisor-rated performance. They have sought confirmation of the claim that people's occupations can be identified from their script.

In 1961 Fluckinger *et al.*¹ reviewed experimental work so far and concluded that evidence "for the relationship of graphological inferences with criteria of interest remained fragmentary". By 1983 Klimoski *et al.*² had this to say: "the general trend of findings is to suggest that

graphology is not a viable assessment method". In 1986 Ben-Shakhar *et al.*³ wrote that "when graphologists base their judgements on spontaneously produced text. . . they can achieve positive, if small, validities. However, when non-graphologists analyse the same data, they achieve similar validities. So does a naive and clearly nonoptimal linear model of the information used in those texts." The most recent study mentioned, in 1991, is almost plaintive: "we have tried our utmost but have failed to provide evidence to support the use of graphology for personnel assessment"⁴.

Why should anyone suppose that calligraphy and character are close-coupled? It is easy to show that there is more to handwriting than a peripheral motor habit. Try writing with your non-preferred hand, or with a pen gripped between your toes or strapped to your elbow, and the result is recognizable as a poor attempt at your normal hand. In some sense, handwriting involves templates or models that are represented at a reasonably high level in the central nervous system.

Chronic conditions such as Parkinsonism or Korsakoff's syndrome have an obvious effect on handwriting, and no doubt an experienced eye can detect the effects of short-term stress. There are also those who claim to be able to detect the sex of a writer from the form his or her script takes — completed examination papers provide a regular anonymous source of experimental material for studies of this kind. And where the initial teaching of handwriting is done to a standard national pattern, as for instance in France, there may be an argument that the development of a personalized script will necessarily reflect the ways in which the personality of the individual departs from the average. Finally, there are of course students of handwriting whose skill at detecting forgery or disguised writing is relied upon by the courts as a matter of evidence.

All of this supports the view that handwriting is personal, is more or less consistently identifiable, and may possibly reflect important facts about its producer. Indeed, handwriting is part of self-presentation. What is dubious is the link between elements of letter formation, slant, size and neatness and supposed personal qualities.

You will not, in the work of graphologists, find quantified judgements of differences among individuals with respect to a fixed set of characteristics. Instead their work is best characterized as epithet plucking. This applicant is honest, hard-working and open to experience, while

that applicant is rather over-anxious but creative and imaginative. This is the method of the literary character sketch, where much is left to the imagination of the reader, and where in any case its subject is fictional. Film adaptations of great novels are often disappointing precisely because much of the characterization is missing or seems a misreading to those who have previously read the book.

Psychologists long ago discovered, in what is pompously known as the Fallacy of Personal Validation (or the Barnum Effect, to give it its everyday name, for it only goes to show that there's a sucker born every minute), the poor critical skills we bring to bear on character descriptions of this kind. There are some descriptions ("not always as extrovert as you sometimes appear"; "usually able to rise to the challenge of problems you meet") which nearly everyone, nearly always, nearly everywhere believes to be true of themselves. If, as part of the sales pitch, you are offered a reading of your own character, do not be surprised at its apparent accuracy and relevance, indeed its apparently privileged access to facts about your character you thought you had kept well hidden. This is a party trick. It involves describing the human condition. There is a long and occasionally entertaining research literature on the topic.

Another popular sales ploy with graphologists is to relate how they detected dishonesty in a sample of handwriting, and — wonder to behold — the writer had at some time in the past been convicted of a criminal offence. The sceptic might wonder what fraction of potential applicants had committed what were in a strict sense crimes, but remained undetected or undiagnosed.

The point is that graphology operates on a naive unreconstructed common-sense view of personality, and so produces accounts that seem readily comprehensible to prospective punters. In doing so, it makes no concessions to scientific methodology or standards, and, as the BPS's statement makes clear, when it is put to the test it fails. That is not to say that all that is needed for effective personnel selection is a sophisticated reconstructed scientific view of personality: psychologists are not without sin in the claims they make. But graphology is a target at which they can safely throw stones. □

S. F. Blinkhorn is at Psychometric Research and Development Ltd, Brewmaster House, The Maltings, St Albans, Hertfordshire AL1 3HT, UK.

1. Fluckinger, F. A., Tripp, C. A. & Weinberg, G. H. *Percept. Mot. Skills* **12**, 67–90 (1961).
2. Klimoski, R. J. & Rafeali, A. J. *occup. Psychol.* **56**, 191–202 (1983).
3. Ben-Shakhar, G., Bar-Hillel, M., Bilu, Y., Ben-Abba, E. & Flug, A. J. *appl. Psychol.* **71**, 645–653 (1986).
4. Cox, J. & Tapsell, J. Pap. pres. *Occupational Psychology Conf. BPS*, Cardiff, 5 January 1991.

* The British Psychological Society, St Andrews House, 48 Princess Road East, Leicester LE1 7DR, UK.

Heroic holograms

D. J. Bishop

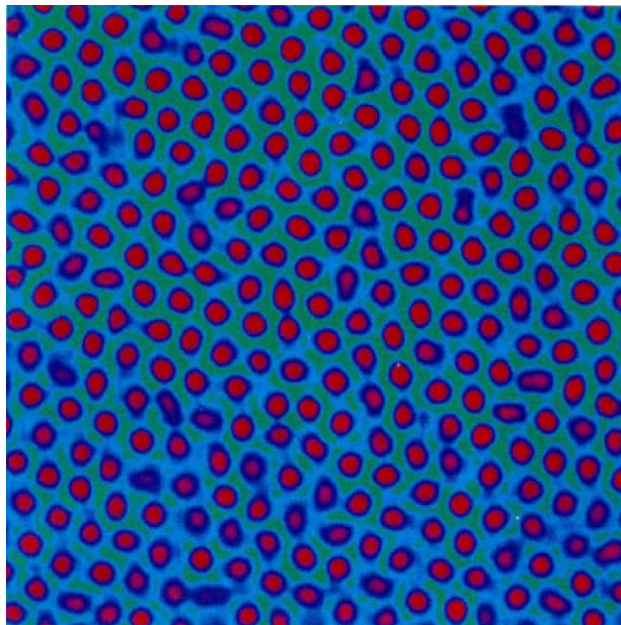
ONE of the pre-eminent intellectual challenges in condensed matter physics is understanding the behaviour of magnetic vortices in the oxide superconductors at the microscopic level¹⁻⁴. In part, this is because the vortices are a nearly ideal model system in which to study the competing effects of disorder, correlation, anisotropy, thermal fluctuations and dimensionality. Another component is the unrequited desire to find commercial applications for these materials, especially in uses that require large critical currents. In the past several years, both experimental⁵⁻¹⁰ and theoretical³ techniques of unprecedented difficulty and sophistication have been brought to bear on the problem. It is a golden era for research in this field.

In 1992, workers at Hitachi developed a technique for imaging magnetic flux lines in a superconductor using electron holography¹⁰, which allowed them to see the behaviour of the lines in real time, as a function of magnetic field, temperature and applied current. In the initial work, they studied niobium; at the time, it was not clear that they would ever be able to apply the technique to the technologically more important but much more difficult high- T_c superconductors. But in this week's issue of *Physical Review Letters*, Harada and colleagues⁵ describe how they have successfully adapted their method to image magnetic vortices in an oxide superconductor, Bi:2212 (formula $\text{Bi}_{2-x}\text{Sr}_{2-x}\text{CaCu}_2\text{O}_{8+\delta}$). It is a truly outstanding piece of experimental physics which moves the bar up another notch for state-of-the-art experiments in this field.

All of the recently discovered oxide superconductors are in the 'type II' class. For these materials, the distance over which magnetic fields can decay (the penetration depth) is much larger than the distance over which the electronic properties such as the superconducting energy gap can change (the coherence length). As was first shown by Abrikosov in the 1950s, conventional type II superconductors can do one of three things when a magnetic field is applied. At low fields and temperatures, they can expel the field (the Meissner state). At high fields and/or temperatures, the field can destroy the superconducting state producing a normal metal. However at intermediate fields (the mixed state), the magnetic field enters the

superconductor in the form of quantized bundles of magnetic flux called magnetic flux lines. In the absence of disorder, these flux lines will form a well-ordered triangular flux lattice.

In the oxide superconductors, experiments have indicated that still another phase can exist — a liquid of flux lines — and that the flux lattice can melt into this flux-line liquid. We care about magnetic



False-colour image of a magnetic flux-line lattice in a crystal of Bi:2212, as seen using magnetic decoration. The flux lines are the red dots and are roughly 1 micrometre apart in the applied field of 20 gauss. Note that the vortices form a triangular lattice.

flux lines because a flowing electric current in a superconductor induces a force on them. If they can respond to this force and are free to move, they will dissipate energy in the flowing current and produce an effective resistance. This is especially true in the vortex liquid state. In this case, these superconductors aren't even good conductors.

It is clear that imaging techniques will be of key importance in unravelling the dynamics of such magnetic flux-line lattices. Those developed over the years include magnetic decoration⁶, STM imaging⁷, small-angle neutron scattering⁸, scanned Hall bars⁹ and the new entrant in the field, electron holography¹⁰. For example, shown in the figure is an image of the flux-line lattice in a crystal of Bi:2212 as seen using magnetic decoration. One can clearly see the individual flux lines and the triangular lattice that they form. But the problem with this technique and most of the others is that they do not allow one to image the lattice and follow its dyna-

mics in real time and real space. The electron holography technique of Harada *et al.*^{5,10} does not suffer from these limitations. Essentially they use very well collimated beams of electrons (divergence angle about 10^{-6} radians) to illuminate an array of vortices. The vortices are then visible in an appropriately defocused image known as a Lorentz micrograph, which is a result of the small phase shifts that electrons undergo when they circumnavigate the magnetic flux lines.

Although simple in concept, the technique is extremely difficult in practice. For example, achieving electron beams that are this well collimated has been the result of many years of work at Hitachi's Advanced Research Laboratory. To do these heroic experiments, which would be impossible in a conventional laboratory environment, they have been forced to build a separate building in which they are isolated from all sources of mechanical, electrical and magnetic noise. These workers will have very little competition.

In this week's paper, Harada and co-workers⁵ report that they have observed vortices in a film of Bi:2212 and have been able to follow them up through the melting temperature of the flux lattice. At low temperatures, they observe a roughly triangular lattice of vortices. As they raise the temperature, they see that at each temperature step the vortices move a little, but eventually settle into a new static configuration. However, at the irreversibility temperature (but well below T_c) they see the lattice fade away. Thus they are the first to

generate real-space, real-time images of a melting magnetic flux-line lattice in a type II superconductor. It is an experimental *tour de force*.

Many of us have been consumed for the past six years by the issue of whether magnetic flux lattices could indeed melt. We have argued for quite some time that they do melt, but seeing is believing — to those in the superconductivity business, these images are nothing less than spectacular. □

David Bishop is at AT&T Bell Laboratories, Murray Hill, New Jersey 70974, USA.

1. Bishop, D. J. *et al. Scient. Am.* **268**, 48–55 (1993).
2. Bishop, D. J. *et al. Science* **255**, 165–172 (1992).
3. Huse, D. A. *et al. Nature* **358**, 553–559 (1992).
4. Bishop, D. J. *Nature* **365**, 394–395 (1993).
5. Harada, K. *et al. Phys. Rev. Lett.* **71**, 3371 (1993).
6. Gammel, P. L. *et al. Phys. Rev. Lett.* **59**, 2592–2595 (1987).
7. Hess, H. *et al. Phys. Rev. Lett.* **62**, 214–217 (1989).
8. Cubitt, R. *et al. Nature* **365**, 407–411 (1993).
9. Hallen, H. D. *et al. Phys. Rev. Lett.* **71**, 3007–3010 (1993).
10. Harada, K. *et al. Nature* **360**, 51–53 (1993).

RÉSUMÉ

Folding network

MYOSIN is known to be involved in the movements of many types of cells other than muscle, but no specific organization of arrays of this molecule into a structure analogous to the sarcomeres of skeletal muscle has been observed in non-muscle cells. Until now, that is, for A. Verkhovsky and G. Borisy (*J. Cell Biol.* **123**, 637–652; 1993), using a specific fluorescent label, have traced the myosin distribution in living fibroblast cells, after removing all other cytoskeletal fibrillar structures (actin, microtubules and intermediate filaments). The authors then examined platinum replicas of the same cells by electron microscopy and discovered a network of myosin minifilaments. Each minifilament contacts its neighbour at the globular head region of the molecule, suggesting that the cell may contract by folding of the myosin filament network.

Blood curdling

THE deerfly, *Chrysops*, is another of those unlovable creatures that evolution has endowed with a biologically active molecule of unusual biochemical interest. Like other bloodsuckers, the insect secretes an inhibitor of clotting, and S. A. Grevelink and co-workers (*Proc. natn. Acad. Sci. U.S.A.* **90**, 9155–9158; 1993) have now found that this is a salivary protein, which eliminates platelet aggregation at minute concentrations. Unlike many common inhibitors of platelet function, the protein does not act by intervening in the second messenger cascade, but rather by competing with fibrinogen at its receptor on the platelet surface. Grevelink *et al.* leave the therapeutic possibilities of the new protein hanging in the air.

Ice and rice

THE ancient methane content of the atmosphere can be reconstructed to some extent from the gas contents of bubbles trapped in the polar ice sheets. How are anthropogenic emissions affecting the record? Profiting from the increased precision of a newly developed ice-milling method to look at the past thousand years' worth of an ice core drilled in central Greenland (the 'Eurocore'), T. Blunier and co-workers have found a strong correlation between methane concentrations and the population of China in the eighteenth century, both of which began to climb steeply after about AD 1750 (*Geophys. Res. Lett.* **20**, 2219–2222; 1993). Wetlands are one key source of methane emission, and the increasing cultivation of rice fields could account for part (although, as the authors are careful to point out, not all) of the increase in concentration. Even before the industrial revolution, man's effect on the methane cycle seems to have been far from negligible.

GEOCHEMICAL WEATHERING

Construction amid destruction

J. Donald Rimstidt

That one might read the book of fate
And see the revolution of the times
Make mountains level, and the Continent,
Weary of solid firmness, melt itself
Into the sea.

(Henry IV Part 2, Act III, Scene I)

ALTHOUGH the destruction of rock by chemical weathering is ever so slow at the macroscopic level, as asserted by Shakespeare, it is dynamic in the extreme at the microscopic level, as recounted by William Casey and colleagues on page 253 of this issue (W. H. Casey, H. R. Westrich, J. F. Banfield, G. Ferruzzi and G. W. Arnold, *Nature* **366**, 253–256; 1993).

The pyroxene and pyroxenoid silicate minerals studied consist of parallel chains of covalently bonded silica tetrahedra decorated with calcium and/or magnesium ions held in place by ionic bonds. Their destruction by chemical weathering proceeds in steps: first, hydrogen ions exchange with the calcium or magnesium cations; next, the cations are leached from the solid.

This leaves behind a leached layer consisting of amorphous silica. The common wisdom amongst geochemists has been that the amorphous layer is quickly destroyed by depolymerization and dissolution. Far from it! Raman spectra obtained by Casey and co-workers show that the silica in the leached layer can react to form four-membered silicate rings. This, together with similar reconstructive processes, seems to stabilize the leached layer to slow its rate of dissolution. Casey and colleagues speculate that, as rocks are weathered, the silica in this reconstructed, leached layer subsequently reacts with cations from the surrounding solution to create new minerals. This model suggests that clay minerals, which are common weathering products, grow through the reconstruction of residual silicate material on weathered minerals rather than by direct precipitation from solution.

Weathering reactions convert minerals to the soil that mantles the continents, and the cations leached from these minerals are carried to the ocean by rivers; both are important steps in the geochemical cycling of elements. Geochemists have long struggled to reconcile the discrepancy between laboratory-measured cation leaching rates and cation fluxes measured in the field, the latter being consistently slower. Could the leached layer that forms on weathered minerals in soils be responsible?

There are two different ways in which it could do this. Leached layers may act as barriers to the attack of hydrogen ions and

thus reduce the rate of the leaching reaction, or they may react with some of the leached cations to form new minerals and thus reduce the apparent rate of leaching. Either effect would explain the slower rate of cation release that is found in field studies, while both effects develop too slowly to be observed in most laboratory studies. In addition, it would explain the apparent topotactic growth of new minerals on the ones being destroyed by weathering (topotaxy is the strongly preferred orientation of crystals produced when a new phase grows on the original phase so as to preserve most of the original bonds). Altogether, this is a compelling explanation.

As with any good scientific investigation, this one creates at least as many interesting questions as it answers. For example, do other silicate minerals show similar evidence of reconstruction during dissolution? And what are the chemical steps that lead to the formation of new minerals from the leached surface? These steps are presumably similar to those that take place during the synthesis of zeolites from the aluminosilicate gels, so if we could understand one of these processes, we would probably find that we had solved the other.

Then there are the biological aspects to consider. How might the silica-rich, leached mineral surfaces interact with lung tissue? Every day we inhale substantial quantities of mineral dusts, some of which remains in our lungs. We already know that some kinds of mineral dust — asbestos minerals inevitably come to mind — cause fibrosis, lung cancer and mesothelioma. Perhaps, however, these diseases are the result of biochemical interactions not with fresh, unleached mineral surfaces but with leached and partly reconstructed surfaces.

Finally, how do these surfaces interact with soil solutions? They might, for instance, sequester plant nutrients so they are not carried away by infiltrating rain water; amorphous silica is widely used in chromatographic separations of organic and biochemical compounds because of its sorptive capacity, and it would be interesting to know how the reconstructed layer on weathering silicate minerals compares to chromatography silica. Can these layers immobilize toxic organic compounds in soils? Casey and co-workers may well find that they have started a flood of new geochemical research. □

J. Donald Rimstidt is in the Department of Geological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA.

Sharper than a needle

Michael Whitaker

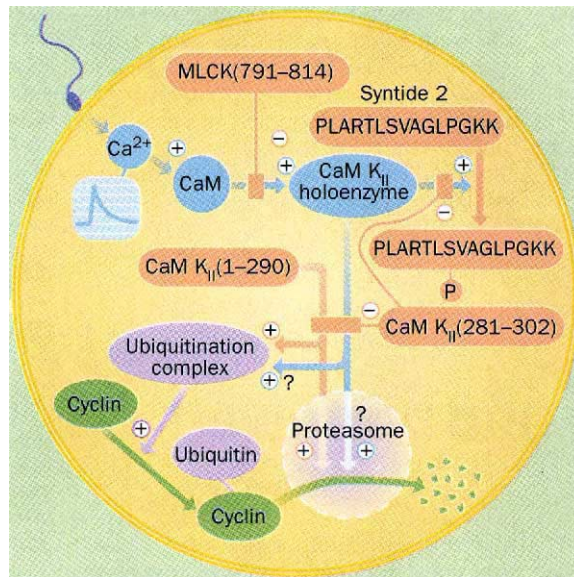
A RECURRING theme in embryology has been finding ways to trick the egg into thinking it is fertilized without recourse to sperm, the most celebrated being Bataillon's activation of the *Xenopus* egg 80 years ago by prodding it with a glass needle. More recently, fertilization has provided a context for the study of cell-cycle regulation. On page 270 of this issue¹, Lorca *et al.* repeat the needle trick, but use far more sophisticated molecular tools to demonstrate that the cell cycle is under the control of the calmodulin-dependent kinase, CaM K_{II}.

When waiting for the sperm, frog eggs sit in metaphase, with chromosomes condensed and aligned on the spindle, ready to move apart. The chromatin and cytoskeleton are maintained in this state by another kinase, MPF (M-phase promoting factor). At fertilization (or after using that needle), an explosive wave of calcium consumes the egg and the cyclin component of the active mitotic kinase is degraded, destroying kinase activity and allowing the egg to move to the next stage of the cell-division cycle.

The link between calcium and proteolysis has now been dissected using a trayful of molecular instruments. The most useful of them is a truncated, unregulated version of the CaM K_{II}, CaM K_{II} (1–290), that is constitutively active and needs neither calcium nor calmodulin. With this molecular equivalent of the glass needle, Lorca *et al.* show that degradation of cyclin and resumption of meiosis result from injecting the unregulated CaM K_{II} (1–290) into the unfertilized egg — this is parthenogenesis at the molecular level. The lesson of Bataillon's needle, though, is that far blunter instruments achieve the same effect. So the purpose of the experiments with the other molecular tools was to see, first whether the CaM K_{II} (1–290) is acting specifically and, second, whether this is the kinase activated by calcium at fertilization.

CaM K_{II} is generous to biologists. As well as the residue 1–290 unregulated catalytic domain, it provides a 281–303 autoinhibitory domain that, as a peptide, will antagonize the native holoenzyme. CaM K_{II} (281–303) blocks cyclin degradation in intact eggs after electric shock, a more modern (though no less crude) version of the needle treatment that induces degradation of cyclin and resumption of meiosis in control eggs. The two CaM K_{II} domains are all that are needed to make

the case for the calmodulin-dependent kinase. It is reassuring, though, that the presence of calmodulin on the signal transduction pathway can also be demonstrated. In extracts of unfertilized eggs, cyclin degradation can be triggered by addition of calcium which, in turn, is blocked by a third molecular tool, a myosin light-chain kinase-derived peptide — MLCK (791–814) — that antagonizes



Pathway of cyclin degradation at fertilization in the *Xenopus* egg. The molecular tools used are highlighted. Observations in ascidians and elsewhere suggest that the proteasome may mediate destruction of the ubiquitinated cyclin.

calmodulin². As might be expected, cyclin degradation by addition of the unregulated CaM K_{II} (1–290) to extracts needs no calcium or calmodulin and is not affected by the MLCK (791–814) calmodulin inhibitor, whereas the autoinhibitory CaM K_{II} (281–303) peptide blocks degradation in response to either calcium or the unregulated enzyme.

Lorca *et al.* employed a fourth molecular probe to satisfy the sceptics who might frown on the use of tools derived from rat brain CaM K_{II} in *Xenopus*. It is syntide 2, a known substrate peptide for the enzyme that has the RXXS/T substrate consensus motif. Syntide 2 is nicely phosphorylated by the unfertilized egg, once it has been given an electric shock to trigger calcium release, and the time course reflects the progress of the calcium activation wave. In this final flourish, the authors show that syntide 2 phosphorylation by activated extracts is blocked by CaM K_{II} (281–302) and thereby demonstrate a native CaM K_{II} activity in the frog.

So neat an application of such precise tools to a problem of cell regulation is

noteworthy in itself. It offers an example of how structural and sequence information gained from molecular biology can generate an experimental strategy, by way of cell physiology, that complements the genetic approaches to which molecular biology more commonly leads. But it also clears up a conundrum, opens up a connection and raises a question.

The conundrum concerns CSF (cytostatic factor), an activity in unfertilized egg cytoplasm that can arrest cells in mitosis by preventing degradation of cyclin. The idea is that CSF keeps the unfertilized *Xenopus* egg waiting in metaphase for the fertilizing sperm by actively preventing the cyclin degradation that would normally follow activation of MPF. CSF has been identified as the product of the *mos* proto-oncogene³. No one knows how *mos* prevents cyclin destruction, but the natural assumption was that calcium would indirectly cause cyclin proteolysis by destroying cyclin's protector, *mos*/CSF (ref. 3).

The odd fact is that degradation of *mos* after fertilization, though triggered by the calcium transient at fertilization, takes much longer than cyclin degradation⁴; this is not (nor has it ever been) consistent with the idea that calcium causes *mos* destruction by calpain⁵, a calcium-activated protease, after which cyclin destruction could occur. Lorca and co-workers had already shown that one half of this mechanism, activation of the calpain protease, could occur only at abnormally high calcium concentrations². They now demon-

strate that *mos* degradation is controlled by CaM K_{II}. What this implies is that although *mos*/CSF is equal to the job of trapping the nucleus in metaphase as the frog oocyte matures, it cannot prevent the fertilization wave of calcium from triggering cyclin destruction and itself later succumbs to the same signal.

The connection is between calmodulin, CaM K_{II} and the machinery that regulates

1. Lorca, T. *et al.* *Nature* **366**, 270–273 (1993).
2. Lorca, T. *et al.* *EMBO J.* **10**, 2087–2093 (1991).
3. Sagata, N., Watanabe, N., Vande Woude, G. F. & Ikawa, Y. *Nature* **342**, 512–518 (1989).
4. Watanabe, N., Hunt, T., Ikawa, Y. & Sagata, N. *Nature* **352**, 247–248 (1991).
5. Watanabe, N., Vande Woude, G. F., Ikawa, Y. & Sagata, N. *Nature* **342**, 505–511 (1989).
6. Hepler, P. K. *Int. Rev. Cytol.* **138**, 239–268 (1992).
7. Lu, K. P. & Means, A. R. *Endocrine Rev.* **14**, 40–58 (1993).
8. Baitinger, C., Alderton, J., Schulman, H. & Steinhardt, R. A. *J. Cell Biol.* **111**, 1763–1773 (1990).
9. Kong, S. K. & Chock, P. B. *J. Biol. Chem.* **267**, 14189–14192 (1992).
10. Shimbra, N. *et al.* *J. Biol. Chem.* **267**, 18100–18109 (1992).
11. Ghisla, M., Udvardy, A. & Mann, C. *Nature* (in the press).
12. Gordon, C., McGurk, G., Dillon, P., Rosen, C. & Hastie, N. D. *Nature* (in the press).
13. Kawahara, H., Sawada, H. & Yokosawa, H. *FEBS Lett.* **310**, 119–122 (1992).

the cell cycle. Calcium transients occur around mitosis in somatic cells as well as embryos, though they are less dramatic than the calcium explosion at fertilization⁶. Overexpression and deletion of the calmodulin and CaM K_{II} genes can alter cell-cycle timing or cause arrest⁷. The selfsame CaM K_{II} (281–302) inhibitory peptide blocks mitosis during embryonic cell cycles⁸. It seems quite possible that the CaM K_{II} will turn out to regulate mitosis quite generally. The mechanism defined in the frog egg may apply and the means to test the principle have now been demonstrated. One admittedly wild prediction from these experiments in the frog is that *mos*/CSF may act by blocking endogenous cell-cycle calcium transients, although, as noted, *mos* is no match for the externally triggered fertilization wave.

The question is what the substrate for the CaM K_{II} might be. Here we have to work backwards from cyclin. Cyclin is degraded by first being tagged with ubiquitin to mark it for destruction. Ubiquitination can be turned on and off by phosphorylation⁹. Proteolysis of the ubi-

quitinated cyclin may be due not, it now appears, to the calpain protease, but to activation of the ATP-dependent proteasome, a heterologous protein-chewing organelle of high relative molecular mass. Antisense disruption of proteasome subunit expression blocks the cell cycle in cell lines¹⁰ and mitotic arrest mutations in proteins homologous to the 26S proteasome subunits have been identified in both fusion and fission yeast^{11,12}. There is evidence, this time from ascidian embryos, that the 26S, ATP-dependent proteasome is active at the metaphase–anaphase transition, at the time when cyclin degradation occurs¹³. Activation of the proteasome seems to be due to conversion from a 20S to 26S form, using addition of a pre-existing subunit (H. Kawahara and H. Yokosawa, personal communication), and is the sort of event that could be controlled by a CaM K_{II} -initiated phosphorylation cascade. Could this be another first for eggs and embryos? □

Michael Whitaker is in the Department of Physiology, University College London, Gower Street, London WC1E 6BT, UK.

SOLAR SYSTEM

Rings old, new and borrowed

Carl D. Murray

THE ubiquity and diversity of planetary ring systems mean that observers have no shortage of targets while theoreticians struggle to find explanations. The situation has now been helped (or hindered, depending on viewpoint) by the possible detection of a new ring system and the imminent prospect of seeing changes to one of the established systems. These and other ring results were reported at a conference held last month in Boulder, Colorado*.

When the fragments of Comet Shoemaker–Levy 9 strike Jupiter next July (see *Nature* 365, 784–785; 1993) the dust particles associated with the comet may not meet the same fate as the individual nuclei, because their orbits differ. One possible outcome for the smaller particles (those less than 2 micrometres across) is a journey through the jovian plasma environment which could lead to the formation of a diffuse ring within 10 years or so (Mihaly Horanyi, University of Colorado). Such a ring is unlikely to be permanent, as its radial location (4.5 to 6 Jupiter radii) would mean that it would be subject to the sweeping action of Io, one of the Galilean satellites.

Jupiter already has a dusty ring, and the prospect of a new influx of material has led

to a re-examination of the relative contributions from such sources as existing dust striking Jupiter's small satellites, continual bombardment by interplanetary micrometeoroids, and the occasional approach of a comet such as Shoemaker–

Levy 9 leading to a “pulse of production” (Doug Hamilton, Cornell University). The results suggest that although the interplanetary flux dominates, cometary events may cause temporary brightening of the ring and produce features that could be longitudinally confined for several weeks. When this new source of dust hits larger bodies in the ring next July, observations of the ring may give clues to the location of undiscovered satellites.

Ring dynamicists are always looking for evidence of small, undetected satellites as these provide a convenient explanation for a variety of ring phenomena. An examination of Voyager 2 images of the faint λ ring of Uranus shows distinct, periodic variations in brightness along its orbit (Mark Showalter, Stanford University). One explanation for the spacing of the features is that they are caused by a perturbing satellite at a radial separation of about 6,700 km from the ring. If so, this satellite must be small enough to have escaped detection by the Voyager cameras, but there is mounting evidence for the existence of such objects in planetary ring systems (M. Showalter, *Nature* 351, 709–713; 1991).

Some planetary rings are eccentric. It is normal for orbits to precess because of the oblateness of the planet, but eccentric rings are observed to precess at a uniform rate despite the fact that particles at the inner and outer edges move on different orbits. The generally accepted theory, proposed by Peter Goldreich and Scott Tremaine in 1979 (*Astr. J.* 84, 1638–1641; 1979), has been that the self-gravity of the ring prevents differential precession. But a recent analysis of Voyager occultation data shows that this theory is incompatible



NASA Science Photo Library

Dusty rainbows — Voyager 2 image of Saturn's rings.

*25th Annual Meeting of the Division for Planetary Sciences, American Astronomical Society, Boulder, Colorado, 18–22 October 1993.

with observations for the ϵ ring of Uranus (Amara Graps, SUNY Stony Brook and NASA Ames). Trying to keep one step ahead, theorists are invoking the effects of viscous instability in the rings to explain this and other discrepancies that have emerged in recent years (Nicole Rappaport, Jet Propulsion Laboratory).

Saturn has examples of the different types of rings seen elsewhere in the Solar System, but it also presents its fair share of problems for dynamicists. "Ring people have become a very desperate lot", admitted Carolyn Porco (University of Arizona). She proposed that some of the unusual structure in Saturn's C and D rings might be caused by resonances set up through acoustic oscillations of the planet. So detailed observations of these rings might provide clues about Saturn's internal structure. The Cassini spacecraft will begin its orbital tour of the saturnian system in 2004 and will help to provide answers to some of the more baffling questions. In the meantime there will be an opportunity to study the extended, diffuse E ring of Saturn in 1995 when the reflected glare of the main rings vanishes as the Earth crosses Saturn's ring plane.

There may be at least one ring system closer to home, with new evidence from both modelling and observations that the Earth is embedded in a ring of asteroidal dust (Stan Dermott, University of Florida). The model shows that as dust spirals in from the asteroid belt towards the Sun under the effects of Poynting–Robertson light drag, it encounters a series of strong resonances close to the Earth where, for example, 20 per cent of particles less than 12 μm across can be trapped. The model also predicts a distinct asymmetry in the leading/trailing distribution of material in this 'solar ring' — and as Dermott found, this agrees with observations made by the Infrared Astronomical Satellite (IRAS) in 1983. The size of the asymmetry should vary as the Earth moves around its orbit, so an analysis of data from the Cosmic Background Explorer (COBE) spacecraft should eventually allow a further test of one of the model's predictions. The dust must be asteroidal, as cometary particles tend to have larger eccentricities and would not get trapped. Although the material eventually escapes from the resonances because its eccentricity increases, there is new dust spiralling in to take its place; the Earth effectively 'borrows' dust from the asteroid belt before returning it to the interplanetary environment.

There is every prospect that the marriage between observation and theory of the abundant rings in the Solar System will continue to be fruitful, despite the unequal contributions of the partners. □

Carl D. Murray is in the Astronomy Unit, Queen Mary and Westfield College, Mile End Road, London E1 4NS, UK.

Chaperonin duet

R. John Ellis

PROTEINS are the action molecules of the cell — it is their precise shapes that ultimately determine the properties of living organisms. In work published on pages 228¹ and 279² of this issue, Ulrich Hartl and colleagues now provide a comprehensive biochemical explanation of one stage of the process by which a protein achieves its individual shape.

Experiments *in vitro* show that the information for the shape of each protein is contained within its sequence of aminoacyl residues, so in principle proteins can assemble themselves spontaneously; most textbooks have it that this is what happens in the living cell. However, we now know that to be expressed efficiently within the complex and highly concentrated milieu inside a cell this information is assisted by pre-existing proteins acting as molecular chaperones^{3–5}. One group of chaperones called the chaperonins assists the correct folding of newly synthesized polypeptide chains, but does not seem to be involved in any subsequent association of folded protein monomers into larger structures. What Hartl and co-workers show is that the mechanism involves the repeated interaction of two chaperonin proteins in an ATP-driven duet; this interaction continues until the polypeptide is folded correctly.

From previous work^{3–5}, it seems that chaperonins help in folding newly synthesized polypeptide chains by preventing the aggregation of partially folded intermediates. Such aggregates form readily *in vitro*, because the intermediates present interactive surfaces to one another before these surfaces are buried in the final structure. Aggregates are dead-ends — they cannot unscramble to form the correct structures. But are these *in vitro* observations relevant to what is going on inside the cell?

Genetic evidence

There is strong genetic evidence that they are, at least in *Escherichia coli*⁶. Experiments with a range of purified proteins show that the chaperonins are not specific for the polypeptides whose refolding they assist *in vitro*, nor do they increase the rate of refolding, only the yield of correctly folded product⁷. The next questions, then, are how it is possible for the chaperonins to assist in the folding of a wide variety of different proteins, and why two components are required.

Chaperonins from the prokaryotic cytosol, mitochondria and plastids come in two forms, chaperonin 60 (or GroEL from *E. coli*) and chaperonin 10 (or GroES from *E. coli*). GroEL consists of two stacked rings of seven identical subunits of

M_r 60K, each ring enclosing a central cavity; GroES consists of a single heptameric ring of identical 10K subunits. In the presence of Mg-ATP or Mg-ADP the GroES ring forms an asymmetric complex by binding to one end of the GroEL cylinder. Each oligomer of GroEL binds one or two molecules of partially folded polypeptide in the end cavities, probably in the form of compact intermediates (collapsed polypeptide chains with some secondary structure but little or no ordered tertiary structure). The cavities are each large enough to accommodate up to a 90K globular protein, and it seems that binding within these cavities provides the essential chaperone function by restricting each molecule of compact intermediate to a sequestered environment where it cannot aggregate with others.

Bond breakage

In addition it has been suggested that this binding causes breakage of noncovalent bonds in the compact intermediate that may otherwise prevent misfolded structures from reaching the correct tertiary conformation⁷; if that is the case, the GroEL acts both as a chaperone to prevent incorrect interactions and as a tutor to correct mistakes. The precise nature of the binding to GroEL is unknown, but is likely to include hydrophobic interactions.

The native tertiary structure forms when the bound compact intermediate is discharged from the internal walls of the GroEL cavities and folds to completion according to the still unknown rules of protein folding, as it does *in vitro* in a classic refolding experiment of the type pioneered by Anfinsen⁸. This type of model has been termed the 'Anfinsen cage model', to emphasize the idea that the chaperonins are not violating the principle of protein self-assembly⁹. This discharge requires Mg-ATP, and the 14 ATPase sites of GroEL act cooperatively¹⁰. Cooperativity of ATP binding and hydrolysis is increased by GroES, which also halves the rate of ATP hydrolysis by inhibiting the active sites in one of the two rings of GroEL⁷. The ATPase activity of GroEL is also greatly increased by the binding of partially folded polypeptide. Treatment of the chaperonin 60 from another bacterium, *Rhodobacter sphaeroides*, with Mg-ATP produces large structural rearrangements that are visible in the electron microscope, and these changes are likely to be a reflection of the discharge process⁹.

Kinetic work⁷ using a fluorescent probe has suggested that ATP binding causes GroEL to cycle between two states which

have alternately high and low affinities for partially folded polypeptides; the cycle continues until the polypeptide reaches its native conformation. This work shed no light on the involvement of GroES, however, because the discharge of the particular polypeptide used is only slightly stimulated by GroES *in vitro*. This lack of an absolute requirement for GroES has been reported for the GroEL-assisted refolding of several polypeptides *in vitro*¹¹. But GroES is both present *in vivo* and essential for cell viability, so it is likely to be involved with GroEL in protein folding in the cell. What, then, is the exact role of GroES?

Reaction cycle

Hartl and colleagues^{1,2} have carried out ingenious *in vitro* experiments that lead to the conclusion that GroEL and GroES are involved in a reaction cycle that continues until the discharged polypeptide has folded to the point where it no longer binds to GroEL; the cycle is driven by the binding and hydrolysis of ATP. The essential feature of this model is the opposing effects of GroES and the partially folded polypeptide on the properties of GroEL; that is, the binding of partially folded polypeptide reduces binding of GroES and vice versa. The authors took advantage of the observation that GroES protects the subunits of only one of the two GroEL rings from removal by proteinase K of 16 carboxy-terminal residues. This truncation does not affect the function of GroEL but does enable the two rings to be distinguished from each other by, for example, crosslinking to labelled nucleotides. These new results, together with other data, are consistent in my view with the following model.

In vivo, with the continual presence of Mg-ATP and GroES, the predominant state is likely to be the asymmetric complex of GroEL with GroES stabilized by ADP bound to the GroEL, as previously suggested⁷. Entry of polypeptide triggers the release of both ADP and GroES to produce a binary complex of GroEL and bound polypeptide. In the absence of ATP this complex is stable, and is the form observed in the discovery of the plastid chaperonin¹². But, *in vivo*, the binding of ATP weakens the interaction between GroEL and bound polypeptide, and induces the binding of GroES to the other ring to produce a ternary complex. Co-operative ATP hydrolysis by both rings then causes release of all parts of the bound polypeptide more or less simultaneously into the central cavity where it folds. If folding is incomplete after ATP hydrolysis, the polypeptide rebinds to GroEL now containing bound ADP and GroES; the cycle repeats until the GroEL-binding motifs in the polypeptide are no longer accessible — that is, until the polypeptide is correctly folded.

So GroES does three things. First, it stabilizes one ring of GroEL in a high-affinity state for binding ADP, this state accepting partially folded polypeptide; the inhibition of ATPase activity in this ring also reduces wasteful hydrolysis of ATP when protein folding is not occurring. Second, it rebinds to GroEL upon exchange of ADP with ATP and increases the cooperativity of subsequent ATP hydrolysis; the discovery that GroES, as well as GroEL, binds ATP² may indicate that GroES donates ATP directly to GroEL as part of its job of increasing the cooperativity of ATP binding. Third, GroES rebinds to the GroEL ring containing bound polypeptide, and so prevents the polypeptide being released into the medium before folding is complete.

Why should ATP binding and hydrolysis be cooperative? Probably it is because the polypeptide is bound in the GroEL cavity at seven or more sites, so preventing premature folding of any part of the chain. But for correct folding it is advantageous for all parts of the chain to be discharged at much the same time, so that the entire primary structure is available for folding at the same instant. Thus the oligomeric structure of GroEL provides not only a sequestered environment for partially folded proteins but also a simultaneous-discharge machinery. Polypeptides whose discharge does not require GroES *in vitro* may bind to GroEL less strongly and/or at fewer sites than those that do. Determination of the crystal structure of GroEL-polypeptide complexes should resolve this point, as well as others such as the connectedness of the two cavities.

This model should appeal to those working on protein refolding who have been discomfited by the discovery of the chaperonins. It suggests that proteins do indeed fold inside cells as they do in experiments *in vitro*, but that they do so in the protected environment of an Anfinsen cage provided by the chaperonin molecular machine. □

R. John Ellis is in the Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK.

1. Martin, J., Mayhew, M., Langer, T. & Hartl, F.-U. *Nature* **366**, 228–242 (1993).
2. Martin, J., Geromanos, S., Tempst, P. & Hartl, F.-U. *Nature* **366**, 279–282 (1993).
3. Gething, M.-J. & Sambrook, J. *Nature* **355**, 33–45 (1992).
4. Hendrick, J. P. & Hartl, F.-U. *A. Rev. Biochem.* **62**, 349–384 (1993).
5. Ellis, R. J. *Curr. Opin. Struct. Biol.* (in the press).
6. Horwich, A. L. et al. *Cell* **74**, 909–917 (1993).
7. Jackson, G. S. et al. *Biochemistry* **32**, 2554–2563 (1993).
8. Anfinsen, C. B. *Science* **181**, 223–230 (1973).
9. Saibil, H. R. et al. *Curr. Biol.* **3**, 265–273 (1993).
10. Gray, T. E. & Fersht, A. R. *FEBS Lett.* **292**, 254–258 (1991).
11. Lorimer, G. H., Todd, M. J. & Viitanen, P. V. *Phil. Trans. R. Soc. B339*, 297–304 (1993).
12. Barraclough, R. & Ellis, R. J. *Biochim. biophys. Acta* **608**, 19–31 (1980).

The magic bullet

THE first antibiotic, Ehrlich's 'Salvarsan', was essentially a poisonous bacterial stain. Like any selective dye, it was taken up specifically by its target cells. It then killed them. This strategy has been recently refined. The coloured drug Photofrin, when injected, spreads throughout the body. But it can be activated locally by aiming a laser at the disease site, generating radicals which attack the local problem without causing trouble elsewhere.

Daedalus now wants to work the same trick with radiation therapy, which at the moment is a woefully indiscriminate weapon. He recalls that heavy elements absorb X-rays strongly — hence the unappetizing 'barium meal' given to a patient to outline his intestines for X-ray examination. A barium-based 'X-ray dye', taken up only by the target cells, could make them selectively sensitive to radiation. A derivative of unstable barium azide, for example, would be broken up into its elements by an incident X-ray photon. The resulting metallic barium would wreck any cell.

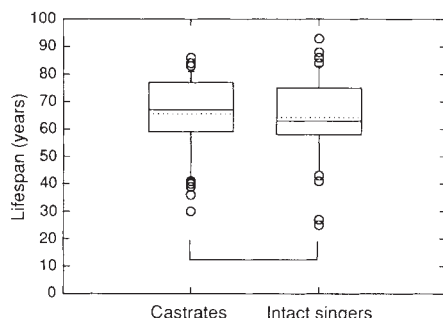
DREADCO's chemists are extending this idea. They point out that most elements, when excited by electron impact, emit specific X-ray fluorescence frequencies. They argue that an element should absorb its own fluorescence frequency extremely strongly, by resonant absorption. So they are devising heavy-element dyes containing more convenient elements than barium, with well-placed X-ray fluorescence frequencies. The end product should be a range of X-ray dyes which, injected into the body, are specifically taken up by the target cells. An intense blast of fluorescence-frequency X-rays, instead of being absorbed evenly by the patient, will dump nearly all its energy into these cells. The ensuing chemical mayhem should kill them instantly.

This trick may work well on parasites, like those of malaria and other tropical diseases, which resist drug therapy. But Daedalus is dreaming up quite a different application. He recalls that high explosive needs a heavy-metal detonator to set it off: typically lead azide. So his new airport-security X-ray scanner works on the lead fluorescence frequency. A detonator will absorb this radiation strongly. Not only will it show up in powerful contrast: the absorbed photons will decompose the unstable azide. An intense, focused blast of the X-rays might even fire the detonator. Passengers could then simply be invited to walk with their luggage through the armoured X-ray scanning tunnel. The survivors could safely be allowed to board their plane.

David Jones

Lifespan and testosterone

SIR — Coronary heart disease and atherosclerosis are the most frequent causes of death among men. Because these conditions occur more frequently in men than in women, it is often assumed that androgens play a causative role. Indeed, correlations between circulating testosterone levels and coronary heart disease have been made, as have negative correlations between blood lipids and androgens, in particular between high-



Lifespan of castrate and intact singers displayed as box plots (matched pairs, 50 in each group). The boxes encompass the 25th–75th percentiles of the data, with the median shown as the solid horizontal line and the mean as the broken line within the box. Whisker caps indicate the 10th and 90th percentile points of the data. Data outside that range are displayed individually. The groups were compared by unpaired *t*-tests and the differences were not significant ($P = 0.65$).

density lipoprotein cholesterol and testosterone¹.

This hypothesis that testosterone may contribute to men's shorter lifespan is difficult to prove experimentally, not least because ethical considerations preclude many obvious tests. But from the sixteenth to the middle of the nineteenth century it was common practice in some European cultures to castrate prepubertal boys to prevent mutation of the voice and thus produce soprano and alto singers. Biographical data of these singers, when compared to those from a control group, should provide information on the influ-

ence of testosterone or — more generally speaking — the presence of testes on longevity.

From encyclopaedias and biographies^{2–9}, we have identified 50 castrates with outstanding reputations as singers born between 1581 and 1858. To establish a control group, we extracted from the same sources the names and dates of 200 'intact' male singers (bass, baritone or tenors) born during the same period and who had achieved comparable fame during their careers. We paired 50 of these singers with the castrates by selecting those with the most closely matched year of birth.

Over 277 years (1581–1858), there was no trend towards a change in lifespan in either group. The mean lifespan of the castrate singers was 65.5 ± 13.8 years (mean $\pm$ s.d.) and of the intact singers 64.3 ± 14.1 years, which is not significantly different (see figure).

These data show that prepubertal re-

moval of the testes had no influence on the longevity of men. But because people living at the time of our sample lived less long, on average, than today, the possibility remains that androgens and other testicular hormones may be of importance during the extra years of life expectancy this century. The latter assumption receives some support from a study reporting a longer lifespan in castrated than in intact mentally disabled patients living in this century¹⁰. But it is difficult to draw conclusions for the general population from a study of institutionalized mentally retarded patients. The possibility remains that the difference in life expectancy between men and women may be related to another factor, probably Y-chromosome-mediated, rather than to testosterone and the testes.

Eberhard Nieschlag

Susan Nieschlag

Hermann M. Behre

Institute of Reproductive Medicine,

Westfalian Wilhelm's University,

Steinfurter Str. 107,

48149 Münster, Germany

Upside-down pendulums

SIR — The inverted pendulum, made stable by rapid vertical oscillations of its pivot, is a well-known curiosity of classical mechanics. It does not seem to be widely known, however, that an inverted double, or even triple, pendulum can be stabilized in the same way, despite theoretical work dating back to 1909 (ref. 1) and a few practical demonstrations^{2,3}. We have performed systematic experiments on these inverted systems and have compared the results with the predictions of a recent theorem⁴.

The pendulums themselves were made from thin-walled stainless steel tubing (3.175-mm diameter), and the couplings had small (7-mm diameter) bearings incorporated which helped to restrict out-of-plane motion while providing only light damping. The length l of each pendulum was 19 cm, so the total length of the triple pendulum, including the intervening bearings, was 57 cm.

The pivot, to which we attached one pendulum on a double-bearing arrangement, was oscillated vertically by an eccentric sliding crank mechanism which provided a reasonable approximation to sinusoidal motion. A powerful d.c. servo-controlled motor provided the drive and the frequency of the applied motion was monitored using an optical shaft encoder.

The drive amplitudes ranged from 1.1 to 1.75 cm, and the drive frequencies ranged up to 45 Hz.

First, we balanced the pendulums in the inverted position using forefinger and thumb. After the motor was switched on, if the frequency was high enough, the

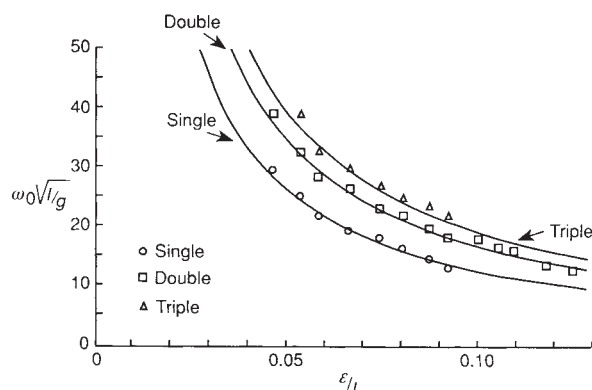


FIG. 1 Comparison between theory (curves) and experiment (points) for the critical value of the drive frequency ω_0 needed for stability of the inverted state. (Some of the higher values of ϵ/l for the double pendulum were attained by using shorter pendulums, with $l = 14$ (rather than 19) cm.)

single and double pendulum inverted states readily became stable. Initial stabilization of the triple pendulum was more troublesome, as out-of-plane motions became very substantial in the high-frequency ranges, but with some practice the simple method described above provided the stable inverted state, which we tested by leaving the system running for several hours.

- Kalin, M. F. & Zumoff, B. *Steroids* **50**, 330–352 (1990).
- Slominsky, N. (ed.) *Baker's Biographical Dictionary of Musicians* (5th edn, Schirmer, New York, 1958; 6th edn, Macmillan, London, 1978).
- Blume, F. *Die Musik in Geschichte und Gegenwart* 4 vols (Bärenreiter, Kassel, 1949–51).
- Brown, J. D. *Biographical Dictionary of Musicians* (Hildesheim, New York, 1970).
- Dean, W. & Knapp, J. M. *Händel's Operas 1704–1726* (Clarendon, Oxford, 1987).
- Ortkemper, H. *Engel wider Willen. Die Welt der Kastraten* (Henschel, Berlin, 1993).
- Riemann, H. *Musik-Lexikon* 10th edn (Hesse, Berlin, 1922).
- Sadie, S. (ed.) *The New Grove Dictionary of Music and Musicians* 1–20 (Macmillan, London, 1980).
- Sadie, S. (ed.) *The New Grove Dictionary of Opera* 1–4 (Macmillan, London, 1992).
- Hamilton, J. B. & Mestler, G. E. *J. Geront.* **24**, 395–411 (1969).

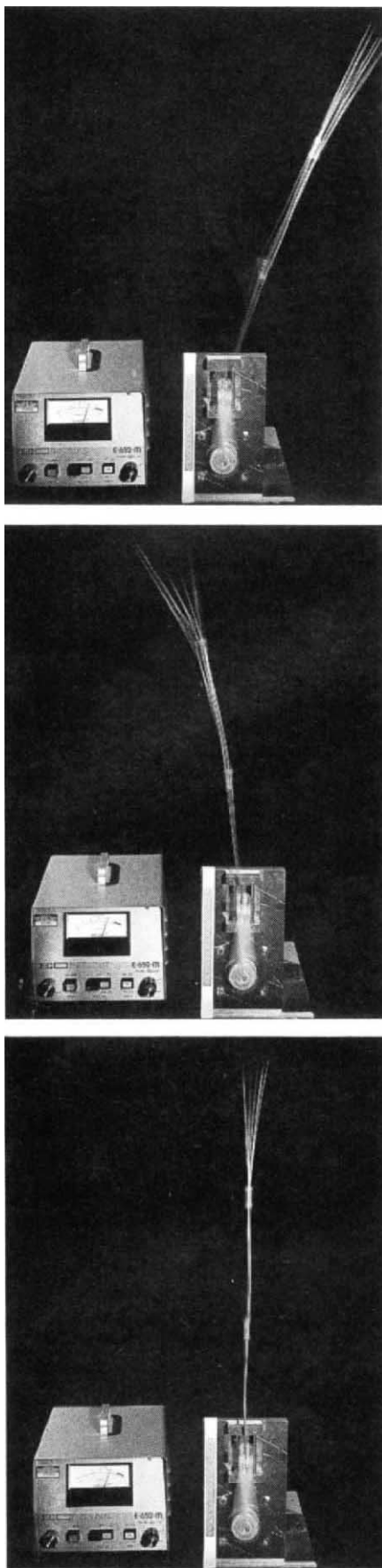


FIG. 2 An inverted triple pendulum returning to the upward vertical after a large initial disturbance. ϵ , 1.4 cm; frequency, ~ 35 Hz.

Once the inverted position had been established, we reduced the frequency of the drive in small steps until the system collapsed. In the case of the single and double pendulums, this was usually preceded by a distinctive lean to one side, which is presumably due to a symmetry-breaking bifurcation close to the limit point of stability. The collapse of the triple pendulum was more sudden and was in some cases accompanied by an oscillation.

Figure 1 shows, for each inverted system, the observed critical drive frequency for various different drive amplitudes ϵ . The theoretical curves are based on the stability condition⁴

$$\omega_0 > \frac{\sqrt{2} g}{\epsilon \omega_{\min}} \quad (1)$$

Here, g denotes acceleration due to gravity, $\epsilon \cos \omega_0 t$ is the imposed vertical displacement of the pivot, and $\omega_{\min}$ denotes the smallest of the natural frequencies of oscillation of the whole system in its undriven, non-inverted (downward-hanging) state. The values of $\omega_{\min}/2\pi$ for our single, double and triple pendulum systems were 1.227, 0.935 and 0.824 Hz, respectively, and these were determined experimentally by exploiting a well-known parametric instability of the non-inverted state⁵ which occurs when ω_0 is increased from zero to twice $\omega_{\min}$. The stability condition (equation (1)) clearly predicts well the observed critical drive frequency for the inverted state. We note, however, that the theory⁴ also predicts the appearance of an oscillatory buckling instability when ϵ/l exceeds 0.091 for the double pendulum or 0.044 for the triple pendulum, yet we saw no clear sign of this comparatively high mode, which may have been damped by frictional effects in the bearings.

We regard the experimental confirmation of equation (1), which is based on linear stability theory, as significant, for one could easily imagine that these inverted pendulums would be unstable to finite disturbances exceeding some very small threshold value. In fact, we find that provided the pendulums are roughly aligned with one another at the start, and equation (1) is satisfied, they will withstand very large initial angular displacements of up to 45° or so, and will gradually wobble back to the upward vertical (Fig. 2).

D. J. Acheson

Jesus College,
Oxford OX1 3DW, UK

T. Mullin

Clarendon Laboratory,
Oxford OX1 3PU, UK

1. Stephenson, A. *Phil. Mag.* **17**, 765–766 (1909).
2. Kaimus, H. P. *Am. J. Phys.* **38**, 874–878 (1970).
3. Chelomei, V. N. *Dokl. Akad. Nauk SSSR* **270**, 62–67 (1983).
4. Acheson, D. J. *Proc. R. Soc. A* **443**, 239–245 (1993).
5. Mullin, T. *The Nature of Chaos* (Oxford University Press, 1993).

Muzzle length and heat loss

SIR — In his interesting News and Views article on the evolution of hominid bipedalism, Wood¹ revives Wheeler's² speculation that elongation of the muzzles of baboons is a thermoregulatory adaptation. According to this suggestion, long muzzles evolved in baboons to provide an expanded surface area of nasal mucosa through which venous blood could flow, thereby cooling the brain with a nasal countercurrent heat exchanger.

Several lines of evidence argue against this interpretation. The muzzles of baboons of the genus *Papio*, unlike those of other savannah-dwelling mammals, contain simple, gently curved nasal conchae that support the nasal mucosa. There is no evidence of an enlargement or an increase in geometric complexity of these surfaces that one might expect if the nasal mucosa were being recruited as a heat exchanger. Nor are there in *Papio* the other conspicuous anatomical adaptations of animals that use a nasal countercurrent heat exchanger: enlarged Steno's glands to provide the water for evaporative cooling from the nasal mucosa or an epiglottic valve to allow the mucosa to be bypassed during rapid thermal panting³. In animals that use the nasal mucosa as an evaporative surface, total ventilation must be increased without increasing alveolar ventilation, as the latter precipitates respiratory alkalosis³. This is accomplished by increasing the dead space in the lungs, a response that does not occur in baboons. Further, with a nasal heat exchanger, baboons would suffer a higher heat loss at all times, including on cold nights, and thus be disadvantaged in the colder or more high-altitude portions of their range.

If muzzle elongation were part of a common adaptive complex for all savannah-dwelling monkeys, one would predict its evolution in non-baboon species occupying hot environments, and that long-nosed species would have lower rates of cutaneous evaporative cooling than short-nosed ones. This prediction is not borne out in, for example, the patas monkey (*Erythrocebus patas*) or the vervet (*Cercopithecus aethiops*). *Theropithecus oswaldi*, a species of large, savannah-dwelling monkey whose remains are found in the same deposits as those of African Plio-Pleistocene fossil hominids, far from showing a trend toward elongation of the muzzle in response to conditions of increasing aridity, showed a progressive reduction in muzzle length over time⁴.

Long muzzles are not restricted to savannah-dwelling *Papio*, but occur in the forest-dwelling mandrill (*Mandrillus sphinx*) and drill (*M. leucophaeus*). All

these forms demonstrate extreme sexual dimorphism in body weight and in the dimensions of the canine teeth. Muzzle elongation (and decoration, in mandrills) appears to have evolved in these animals as part of the "anatomy of bluff and display"⁵, as long muzzles facilitate the flaunting of the males' long, dagger-like canines during open-mouthed gape displays and may also help to attract mates. In male *Papio*, muzzle growth and canine eruption continue after the animals' molars are fully erupted⁶ and may be under hormonal control. Muzzle elongation thus appears to have evolved *pari passu* with increased sexual dimorphism in both body weight and the canine display mechanism.

Nina G. Jablonski

Department of Anatomy and Human Biology, and Centre for Human Biology, The University of Western Australia, Nedlands, Western Australia, 6009 Australia

1. Wood, B. *Nature* **363**, 587–588 (1993).
2. Wheeler, P. *New Scientist*, No. 1612, 62–65 (1988).
3. Taylor, C. R. *Int. Rev. Physiol. Envir. Physiol.* **115**, 119–146 (1977).
4. Jablonski, N. G. (ed.) in *Theropithecus: The Rise and Fall of a Primate Genus* (Cambridge University Press, 1993).
5. Washburn, S. L. & Ciochon, R. L. *Am. Anthropol.* **76**, 96–98 (1976).
6. Freedman, L. *Ann. Transv. Mus.* **23**, 122–262 (1957).

Viral extinctions in deep-sea species

SIR — Coccolithophoridae and planktic Foraminifera have left, in deep-sea sediments, an abundant record that is continuous through time and affords a means to study evolution in progress. Evolution in a population isolated in a marginal marine environment can create a new species in less than 10,000 years¹, leaving hardly any record of the intermediate steps. In contrast, phyletic evolution, also observed in the deep-sea record, proceeds at rates that range from 100,000 years to millions of years and leaves an abundance of intermediate forms (the famous 'missing links')^{1–3}. New species are rapidly distributed across the ocean by surface currents, and are seen to appear in the record abruptly and over wide areas.

The deep-sea record includes episodes of mass extinction, the most conspicuous of which is the one that marks the end of the Mesozoic when most species of Coccolithophoridae and planktic Foraminifera were exterminated. Mass extinctions are rare events that characteristically affect a variety of taxa in a variety of environments — coccolithophorids to dinosaurs. Most commonly, however, species disappear individually (the so-called background extinctions) leaving the neighbouring species undisturbed. An environmental upset is thus excluded, and so is the arrival of a 'better' species. In fact, the

SPECIES EVENTS IN THE PAST 2 MILLION YEARS

Species	Taxon	Appearance	Time
<i>Globorotalia flexuosa</i>	F	Last	68,000
<i>Globoquadrina pseudofoliata</i>	F	Last	220,000
<i>Emiliania huxleyi</i>	C	First	275,000
<i>Globorotalia flexuosa</i>	F	First	401,000
<i>Globorotalia hirsuta</i>	F	First	450,000
<i>Pseudoemiliania lacunosa</i>	C	Last	460,000
<i>Globorotalia tosaensis</i>	F	Last	650,000
<i>Pulleniatina finalis</i>	F	Last	1,400,000
<i>Globigerinoides fistulosa</i>	F	Last	1,560,000
<i>Calcidiscus macintyre</i>	C	Last	1,590,000
<i>Gephyrocapsa oceanica</i>	C	First	1,680,000
<i>Gephyrocapsa caribbeanica</i>	C	First	1,740,000
<i>Discoaster brouweri</i>	C	Last	1,950,000
<i>Globorotalia exilis</i>	F	Last	2,100,000

C, coccolithophorid; F, foraminifer. Time in years before present. (Data from ref. 16 and from W. A. Berggren and M. Berggren (personal communication)).

deep-sea record shows that appearances and disappearances are usually unrelated, causing diversity to change up and down by a factor of three during the Cenozoic (the past 65 million years). The table lists the events that occurred during the past 2 million years. Similar events are exhibited by other marine planktic taxa that also left an abundant record in deep-sea sediments — the radiolaria, diatoms and silicoflagellates. No convincing explanation for background extinction has ever been advanced.

As soon as the first marine virus was reported⁴, I speculated⁵ that background extinctions of abundant and widespread marine species are caused by viruses, the only agents that are species-specific and can thus lead a species to extinction without affecting its neighbours. For such an extinction to occur, however, not only must a suitable virus appear somewhere in the world ocean, but the host species must also be abundant (to facilitate contact) and well mixed (to facilitate spreading).

My original hypothesis has received support from the discovery that viruses are in fact plentiful in the ocean^{6–9}, and from the additional discovery that viruses, rather than nutrients, constrain the blooms of phytoplankton¹⁰. A striking example is the bloom pattern of the coccolithophorid *Emiliania huxleyi* in the North Atlantic^{11,12} (for a picture of a giant bloom from outer space, see Fig. 6.4, p. 145, in ref. 12). The genome of an abundant and widespread species is not uniform, of course, but viral genomes can be expected to be even less uniform because of their high rates of mutation. As a result, the threat to the species remains until the species disappears or the viral genome mutates radically.

The extinction of an established, abundant species vacates a niche that may be occupied by a suitable mutant from a marginal environment if such a mutant is available. If not, diversity decreases. In this model, named 'extinctive evolution'⁵, extinction precedes replacement and the

replacement (if any) need not be better adapted than its predecessor. In contrast, both the standard model (Huxley's "modern synthesis" and updates) and phyletic evolution predict that the new species should be better adapted to the specific environment, or more viable in it, than its predecessor.

The most abundant coccolithophorid species alive today is *E. huxleyi*, illustrated on the cover of the 9 April 1992 issue of *Nature*. The density of *E. huxleyi* in the top 200 m of the ocean averages about 1,000 individuals per litre¹³, yielding a standing crop of 7×10^{22} living cells. *Homo sapiens* is the most abundant primate species. There are 5×10^9 people on Earth, each one with about 10^{13} cells, yielding a standing crop of 5×10^{22} living human cells. Both *E. huxleyi* and *H. sapiens* came into existence at about the same time, about 300,000 years ago, if one includes the *praesapiens* type. Both species are abundant, widespread and well mixed. The numbers of living cells of the two species are similar — close to one 'avogadro' each. (I have introduced the 'avogadro' (Av) as a dimensionless unit of numerical magnitude equal to g/u , where

1. Wei, K.-Y. & Kennett, J. P. *Paleobiology* **14**, 345–363 (1988).
2. Malmgren, B. A. & Kennett, J. P. *Paleobiology* **7**, 230–240 (1981).
3. Malmgren, B. A. *et al.* *Paleobiology* **9**, 377–389 (1983).
4. Mayer, J. A. & Taylor, F. J. R. *Nature* **281**, 299–301 (1979).
5. Emiliani, C. *J. theor. Biol.* **97**, 13–33 (1982).
6. Proctor, L. M. & Fuhrman, J. A. *Nature* **343**, 60–62 (1990).
7. Børsheim, Y., Bratbak, G. & Heldal, H. *Appl. envir. Microbiol.* **56**, 352–366 (1990).
8. Cottrell, M. T. & Suttle, C. A. *Mar. Ecol. Progr. Ser.* **78**, 1–9 (1991).
9. Heldal, M. & Bratbak, G. *Mar. Ecol. Progr. Ser.* **72**, 205–212 (1991).
10. Suttle, C. A., Chan, A. M. & Cottrell, M. T. *Nature* **347**, 467–469 (1990).
11. Bratbak, G., Egge, J. K. & Heldal, M. *Marine Ecol. Progr. Ser.* **93**, 39–48 (1993).
12. Westbroek, P. *Life as a Geological Force* (Norton, New York, 1991).
13. Okada, H. & Honjo, S. *Deep-Sea Res.* **20**, 355–374 (1973).
14. Emiliani, C. *J. geol. Educ.* **39**, 31–33 (1991).
15. Emiliani, C. *Planet Earth* (Cambridge Univ. Press, 1992).
16. Berggren, W. A., Kent, D. V. & van Couvering, J. A. *Geol. Soc. Mem.* **10**, 211–250 (1985).

g is gram and *u* is atomic mass unit, and redefined 'mole' as the mass of one 'avogadro' of items¹⁴.) Extinctive evolution predicts that when a genome becomes as abundant, widespread, and well mixed as *E. huxleyi* and *H. sapiens* are today, viral pandemics are likely (see p. 552 of ref. 15). Indeed, both species are already under viral attack. What remains to be seen is which one of the two will become extinct first.

Cesare Emiliani

Department of Geological Sciences,
University of Miami,
Coral Gables, Florida 33124, USA

Fe₃O₄ and Fe₃S₄ in a bacterium

SIR — Magnetotactic bacteria, a diverse group of Gram-negative prokaryotes that orient along geomagnetic field lines¹, contain magnetosomes, membrane-bounded intracellular single-magnetic-domain particles of either an iron oxide, magnetite (Fe₃O₄) (ref. 2), or an iron sulphide, greigite (Fe₃S₄) (ref. 3). Particles from any given magnetotactic bacterial species or strain are of a specific mineral type with specific crystallographic parameters^{2,3}. Thus, the biomineralization of the magnetosome mineral phase is a highly regulated process^{2,3} probably directed and controlled at the gene level.

We are investigating magnetotactic bacteria from the Pettaquamscutt Estuary (Rhode Island), a chemically stratified semi-anaerobic basin⁴, where the oxic-anoxic transition zone (OATZ) is in the water column because of the upward diffusion of hydrogen sulphide generated from bacterial sulphate reduction. At least seven different morphologically distinct magnetotactic forms exist in 'plates' together with other bacterial types at or close to the OATZ. Populations of one of these, a morphologically distinct, elliptically shaped organism (about 3.1 × 1.3 μm) were as high as 2 × 10⁵ cells ml⁻¹ at

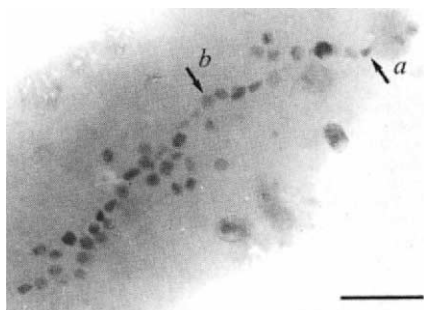


FIG. 1 Electron micrograph of a magnetosome chain within a single magnetotactic bacterial cell containing particles of arrowhead-shaped (a) and rectangular (b) morphologies. Arrows, particles from which the electron diffraction patterns shown in Fig. 2 were obtained. Scale bar, 500 nm.

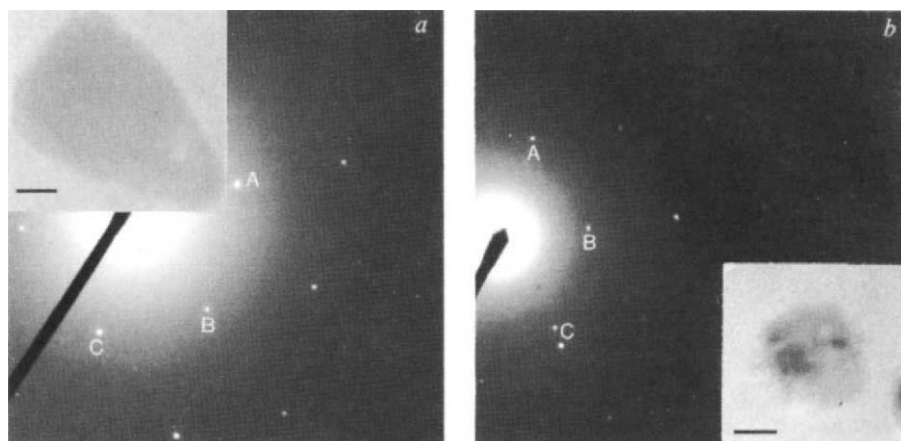


FIG. 2 Single-crystal electron diffraction patterns recorded from the arrowhead-shaped and rectangular particles in the chain shown in Fig. 1. Insets, crystals in the orientation corresponding to the electron diffraction patterns. Scale bar in insets: a, 10 nm; b, 30 nm. For the electron diffraction patterns, the camera length and wavelength were 1,025 nm and 0.02508 Å, respectively. a, The pattern corresponds to the [110] zone of magnetite, Fe₃O₄. Reflection A (220) (2.96 Å); reflection B (113) (2.53 Å); reflection C (113) (2.53 Å). Angles 220 × 113 = 64.8°; 113 × 113 = 50.5°; 220 × 113 = 115.2°. Space group *Fd3m* (cubic); *a* = 8.396 Å. b, The pattern corresponds to the [011] zone of greigite, Fe₃S₄. Reflection A (311) (2.98 Å); reflection B (022) (3.49 Å); reflection C (311) (2.98 Å). Angles 022 × 311 = 64.8°; 311 × 311 = 50.5°; 022 × 311 = 64.8°. Space group *Fd3m* (cubic); *a* = 9.876 Å.

the OATZ and also existed in the anoxic zone below the OATZ. Cells of this organism usually contained a double chain of an average of 37 magnetosomes. Particles within the same chain had two different morphologies: arrowhead-shaped (about 76 × 38 nm); and roughly rectangular (about 64 × 49 nm) (a and b in Fig. 1, respectively).

We used energy dispersive X-ray analyses to show that arrowhead-shaped and rectangular particles consist of iron and oxygen and iron and sulphur, respectively. Single-crystal electron diffraction patterns of the arrowhead-shaped and rectangular particles (Fig. 2) are consistent with the face-centred-cubic spinel minerals magnetite and greigite, respectively. The magnetite particles were elongated along the <111> crystallographic axis and terminated at one end with a well-defined {111} face, whereas the greigite particles were elongated along the <100> axis and terminated with two well-defined {100} faces.

The organism described here is unique among magnetotactic bacteria in that it contains both magnetite and greigite in the same cell. This finding is important to the understanding of the biomineralization process(es) that occur in magnetotactic bacteria. Both crystal types are positioned within the same chain(s), are mineralized with consistently specific sizes and crystallographic parameters, and have their long axes oriented along the chain direction. Because the two crystal types have different crystallographic parameters that have been observed for particles in other magnetotactic bacteria with single component chains^{3,5}, and different crystallographic orientations with respect to the chain, it is likely that

they are produced by two separate biomineralization processes⁶ controlled by separate sets of genes. The organism described here seems to have both sets of genes, whereas other magnetotactic bacteria have only one set.

The large numbers of the bacterium described here could have a significant impact on iron cycling at the OATZ and in the anoxic zone, and could contribute to the deposition of both fine-grained magnetite and greigite that, in turn, contribute to the remanent magnetization of sediments.

Dennis A. Bazylinski

Marine Science Center,
Northeastern University, East Point,
Nahant, Massachusetts 01908, USA

Brigid R. Heywood

Department of Chemistry and Applied
Chemistry,
University of Salford, Salford M5 4WT, UK

Stephen Mann

School of Chemistry,
University of Bath, Bath BA2 7AY, UK

Richard B. Frankel

Department of Physics,
California Polytechnic State University,
San Luis Obispo, California 93407, USA

1. Blakemore, R. P., Blakemore, N. A., Bazylinski, D. A. & Moench, T. T. *Bergey's Manual of Systematic Bacteriology* Vol. 3 (eds Staley, J. T. et al.) 1882–1889 (Williams and Wilkins, Baltimore, 1989).
2. Meldrum, F. C., Heywood, B. R., Mann, S., Frankel, R. B. & Bazylinski, D. A. *Proc. R. Soc. B* **251**, 231–236; 237–242 (1993).
3. Heywood, B. R., Bazylinski, D. A., Garratt-Reed, A. J., Mann, S. & Frankel, R. B. *Naturwissenschaften* **77**, 536–538 (1991).
4. Stolz, J. F. in *Biomineralization Processes of Iron and Manganese—Modern and Ancient Environments* (eds Skinner, H. W. C. & Fitzpatrick, R. W.) 133–145 (Catena, Cremlingen-Destedt, 1992).
5. Mann, S., Sparks, N. H. C. & Blakemore, R. P. *Proc. R. Soc. B* **231**, 469–476 (1987).
6. DeLong, E. F., Frankel, R. B. & Bazylinski, D. A. *Science* **259**, 803–806 (1993).

Politics caught unawares

John de la Mothe

The Sputnik Challenge: Eisenhower's Response to the Soviet Satellite. By Robert A. Divine. Oxford University Press: 1993. Pp. 245. \$25, £19.95.

In the years immediately after the Second World War, there was little doubt in most people's minds that the United States had emerged as the undisputed world leader in science and technology. The brute power unleashed by the Manhattan Project, coupled with the rhetorical force of such statements as Vannevar Bush's *Science: The Endless Frontier*, firmly secured that perception.

The United States was confident that it faced no real scientific, technological or security threat from any other nation. This was partly confirmed — at least to the satisfaction of the White House — by the surveillance reports from flights of U-2 spy-planes over the Soviet Union, which began in 1954. Yet pressure remained from within the Pentagon and the scientific community to launch a satellite programme as quickly as possible: policy makers knew that a diplomatic scandal would erupt if it was discovered that the United States was regularly sending spy-planes over allied territory during peacetime; and the technology needed to put a satellite into space would also be capable of launching intercontinental and intermediate-range ballistic missiles.

In 1955, the White House, anxious to preserve a peaceful image, announced the launch of a series of "small man-made earth circling satellites" — Project Vanguard, as it was called. The proposal was warmly received internationally: satellites were expected to aid worldwide radio communications, as well as other peaceful cooperative activities.

In December 1956 and May 1957, the last two US Viking vehicles were wheeled out to serve as Vanguard test vehicles. The tests were flawless, with the final Viking making a beautiful night ascent. US scientists and policy makers were pleased with their progress. So confident was the White House that it repeatedly led the American people to believe that they would be the first into space and that they would have the first satellite.

But on 4 October 1957, the American dream of technological superiority was suddenly shattered when the Soviet Union launched into space a 184-pound, 22-inch-diameter metal ball. Composed of black painted aluminium panels, Sputnik 1 orbited at up to 18,000 miles per hour some 550 miles above the Earth. The Soviets were thus the first nation into space, the first to launch a man-made satellite and — unbelievably — the first to have the potential to launch inter-

continental and intermediate-range ballistic missiles. The US trauma was almost immediately compounded on 3 November when the Soviet Union launched the much larger Sputnik 2, weighing 1,121 pounds and carrying the dog Laika, the first living animal in space. The third, and by no means final, wound came on 6 November when the first complete Vanguard exploded in a ball of flames less than one second after lift-off.

Sputnik as "an ominous event". The *New Republic* feared that Sputnik was "proof of the fact that the Soviet Union [had] gained a commanding lead in certain vital sectors for world scientific and technological supremacy". Democrats called Sputnik a "devastating blow" and "proof of a growing worldwide Communist supremacy". Such partisan charges led Republicans to play down the significance of the launch by calling it "a silly bauble".

On his return to Washington, Eisenhower was inundated by visits, telexes and telephone calls from worried advisers across the country; totally failing to appreciate the anxious mood of the American people, he decided to adopt a subdued stance. After all, his intelligence reports suggested that the Soviets were

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

High hopes — a group watches as Sputnik 1 passes over Moscow on 12 October 1957.

The launch of Sputnik stunned US scientists and politicians; after two anxious days of waiting for confirmation of the flight, the president of the US National Academy of Sciences sent a feeble letter of congratulations to the head of the Soviet Academy.

White House staff worked feverishly to minimize the damage of the news. But President Dwight Eisenhower, who had gone golfing in Gettysburg the day before the launch, let Press Secretary James Hagerty and Secretary of State John Dulles handle the administration's initial response. Between them they agreed to say that the Soviet achievement came as "no surprise" and to avoid re-asserting their civilian plans for space in order "not to appear scared".

The public reaction was markedly different. Edward Teller, "the father of the hydrogen bomb", screeched that the United States "had lost a battle more important than Pearl Harbour". Magazines reacted by publishing shrill articles asking whether "America Is Going Soft". *Time* and *Newsweek* saw

not a threat. He speeded up plans for the initial deployment overseas of intermediate-range ballistic missiles. An agreement was signed by Eisenhower and the British Prime Minister, Harold Macmillan, to station missiles in British bases, within striking distance of Moscow.

At the same time, Eisenhower was pressed by White House staff and Pentagon officials publicly to use U-2 spy-plane data, which had revealed rocket testing in remote areas of the Soviet Union, but he decided resolutely not to. He also decided to protect absolutely the integrity of the U-2 project. To many, this was the wisest and most courageous decision of Eisenhower's presidential career. As Barry Goldwater said: "Ike took a lot of heat, and kept his mouth shut."

Sensing the public's unease, Eisenhower decided to give a series of news conferences and radio talks. These 'chin-up' chats were "designed to instill confidence and defuse the apparent psychological advantage" that the Soviets had taken. In some of these he was deliberately evasive and misleading; neverthe-

Wide World Photos

less, his low-key response utterly failed, and the passion play continued to unfold. On 31 January 1958, America's first satellite was launched from a modified Jupiter. The satellite, Explorer 1, weighed 10.5 pounds. But any sense of relief was quashed on 15 May 1958 when the Soviets launched Sputnik 3, an immense 7,000-pound 'flying laboratory'. America's first Vanguard finally got off the ground on 17 March and it holds the distinction of having been the first satellite to use solar cells and, still in orbit today, is designed to stay there for a thousand years. Vanguard satellites 2 and 3, launched on 17 February and 18 September 1959, brought the programme to an end. Like Vanguard 1, they too are still in orbit.

Robert Divine provides a fascinating look at the early space race, and in particular at Eisenhower's handling of the Sputnik calamity. He vividly recreates the national hysteria over the first two Sputnik launches, indicating the nervous hand-wringing that the Democrats aggressively exploited for political gain. Divine takes us to private White House meetings, showing how Eisenhower worked closely with James Killian, allowing him to take the lead in creating a civilian agency — NASA (National Aeronautics and Space Administration) — which provided intelligent and forceful leadership for the US space programme. But the president also knew from the U-2 flights over the Soviet Union that the United States had little to fear from the publicly touted 'missile gap', and he quietly fought to limit the number of military missile programmes. Eisenhower's public assurances rested on classified information; his inability to convince the public was ultimately his political undoing.

To many, Sputnik brought an entire way of life into question. The vitality of US education, science and political leadership was thrown into the arena of public debate. The Soviet launch created a crisis of public confidence, and opened a debate on education, science, space exploration and national security that lasted well into the 1960s. Sputnik contributed greatly to the election of President John F. Kennedy in 1960, to the passage of massive federal aid-to-education measures under another Democrat president, Lyndon B. Johnson, to the decision to send men to the Moon and to the Cold War. All because of a 'silly bauble'. All because politics had been caught unawares. □

John de la Mothe is in the Faculty of Administration, University of Ottawa, Vanier Hall, 275 Nicholas Street, Ottawa, Canada K1N 6N5.

Autumn Books

Nature's Autumn Books supplement will be published next week.

How to build an atomic bomb

A. P. French

Critical Assembly: A Technical History of Los Alamos During the Oppenheimer Years, 1943–1945. By Lillian Hoddeson, Paul W. Henriksen, Roger A. Meade and Catherine L. Westfall. Cambridge University Press: 1993. Pp. 509. £45, \$39.95.

THE creation of the first atomic bombs must surely rank as the most spectacular (and dire) marriage of science and technology in the twentieth century, and perhaps in all of history. From the discovery of nuclear fission in late 1938 to the production of the first nuclear explosion in July 1945 was a matter of little over six

was brought into being in April 1943; the general outlines of the project had already been established. The basic theoretical aspects of bomb design had been discussed at a meeting at Berkeley chaired by J. R. Oppenheimer in July 1942; microgram amounts of separated uranium isotopes and plutonium for cross-section measurements had been produced; and techniques for large-scale uranium and plutonium separation were being actively developed. The initial focus was on a gun assembly in which two subcritical masses of uranium or plutonium would be rapidly assembled into a supercritical mass by ordinary explosives, although the possibilities of implosion were also being explored and Edward Teller was already vigorously promoting the notion of a super-bomb using deuterium surrounding a fission-bomb core.

Los Alamos picked up from there. Almost everything of practical importance remained to be done, from learning a great deal more about the nuclear parameters (cross-sections for fission, neutron scattering and neutron capture at all energies in various materials) to developing a reliable gun design, researching the chemistry and metallurgy of uranium and plutonium, and exploring the feasibility of implosion — plus, of course, waiting for the huge production facilities at Clinton, Tennessee, and Hanford, Washington, to come through with kilogram amounts of plutonium and separated ²³⁵U.

Institute for Advanced Study

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Oppenheimer — academic and brilliant leader.

years, and it was achieved against all the odds, and in the face of almost incredible technical difficulties. To conceive of making weapons from one type of atom (uranium-235) that would need to be separated from its chemically identical sister, 140 times more abundant, and another type of atom (plutonium-239) that did not even exist on Earth, was the stuff of science fiction. *Critical Assembly* describes, in great detail, what was involved in making this fantastic proposal succeed, and in particular what took place at Los Alamos, which became the headquarters of the project and the place where the nuclear bombs themselves were developed and assembled.

The United States was fully committed to an atomic bomb programme about 18 months before the Los Alamos laboratory

The unlikely combination of General Leslie Groves, the career officer, and Oppenheimer, the academic physicist, worked remarkably well at Los Alamos, although, as the book points out, the question of whether it would be a military or a civilian establishment was never formally resolved. The account makes clear that Oppenheimer was a brilliant leader of the laboratory. Although a theorist, he had a keen understanding of the problems of experimentalists, and his administrative judgements were most impressive. Nowhere does this show up more clearly than in his willingness and ability to reorganize and reallocate the human resources of the laboratory in response to changing needs. The prime example of this is probably the implosion programme. Even though implosion was recognized early on as a good possibility in principle, a gun assembly promised to be much more straightforward, and it was initially assumed that this could be used for both

uranium and plutonium. But then it was realized that the intense alpha-particle emission of the relatively short-lived plutonium, acting on minute amounts of impurities, could produce neutrons leading to premature initiation of the fission process — and hence a “fizzle” — in a gun system. The high spontaneous fission rate of reactor-produced plutonium (due to ^{240}Pu and not discovered until July 1944) exacerbated the problem and forced a major reorientation of the programme.

In August 1944, Oppenheimer proceeded to focus the main efforts of the laboratory on implosion systems using plutonium, because the necessary separation of ^{235}U from ^{238}U was proving to be slow and difficult, whereas the reactor production of plutonium, and its separation from its chemically different parent (^{238}U) and from fission products, promised to be relatively straightforward. Plutonium thus became the favoured material; but to use it in a weapon required the development of sophisticated two-component lenses made of conventional explosives to drive the implosion, plus perfect synchronization of the multiple detonators for these lenses arranged around the fissile core. These matters became a priority from then until the end of the war.

The story so briefly touched on above has of course been the subject of many books, including three official American histories: Henry DeWolf Smyth's *Atomic Energy for Military Purposes* (1945), Richard Hewlett and Oscar Anderson's *The New World, 1939/1946* (1962) and David Hawkins's *Project Y: The Los Alamos Story* (1961, 1983). *Critical Assembly*, although not strictly an official history, is an officially sponsored one, written under the auspices of the Los Alamos Laboratory and copyrighted by the US Department of Energy. (The authors are at pains to point out that, although certain technical facts remain classified, they were subject to no other constraints.) This book differs from Hawkins's account in several respects, identified by the authors as follows: “it is a history of the technical developments; it is based on the full complement of documents, both classified and unclassified, of wartime Los Alamos; and it explores for the first time the methodology by which researchers at Los Alamos succeeded in their wartime mission.” Also, to quote again from the preface, the authors have tried “to explain how the developments were shaped by individuals and how they related to earlier efforts” — for example, to prior explosives research in England. And, as the authors say, they had the advantage of writing four decades later, thus benefiting from fuller information and an improved historical perspective. I do not think, however, that the book contains any major new revelations.

With its emphasis on being an objective technical history, the book is very short on colourful details or background concerning either people or events. (Where such details do occur, most often in the footnotes, they have a startling effect: for example, one individual is described as having been “slapdash” and another as “ambitious, publicity-prone”; and we learn that a ski slope for the community was cleared of trees with laboratory explosives.) The book will, I think, be of greatest interest to other historical workers. It does show signs of the extended and laborious effort involved in its preparation. The word “recently”, for example, appears in a number of places where the associated references are to interviews at various times between 1984 and 1988.

It is frustrating that, for all the density of narrative detail, a reader is unlikely to end up with a much clearer picture of what

these first atomic bombs were really like. The only diagram in the whole book (as distinct from relatively uninformative photographs) is a crude sketch from *The Los Alamos Primer* (Robert Serber, University of California Press, 1992) showing an imagined implosion arrangement from the earliest days of the project. Perhaps this is because historians do not think in terms of diagrams; or perhaps it was a consequence of continuing secrecy restrictions. I concede, however, that this book fulfils its declared intention of providing an unprecedentedly full picture of all the actions and processes that were involved in bringing the bombs into existence. □

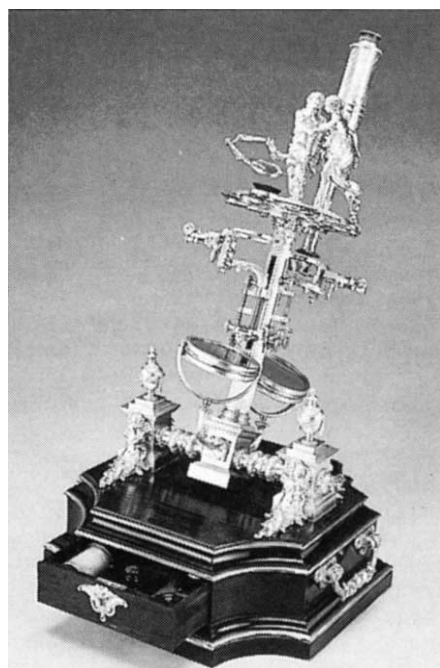
A. P. French, a junior member of the British group at Los Alamos, 1944–46, is in the Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA.

Instruments fit for a king

Willem Hackmann

Public & Private Science: The King George III Collection. By Alan Q. Morton and Jane A. Wess. Oxford University Press: 1993. Pp. 710. £55.

ON Monday 15 November the famous George III Collection of scientific instruments was formally opened at the Science Museum in London. To celebrate this remarkable collection there is a lavish book-cum-catalogue, which is the result of years of painstaking work by its authors.



Silver microscope (c. 1763) — one of a pair made by George Adams for George III and his brother.

They have carefully pieced together the collection from the manuscript sources, and have related the instruments to descriptions in contemporary textbooks and learned papers.

The book is in two sections. The instruments are put in their historical context in the first section, and a five-part catalogue makes up the second. The collection is an amalgam of two main groups of instruments: the equipment used by Dr Stephen Demainbray, an itinerant lecturer who briefly taught science to the Prince of Wales (the future George III) and his brother Edward, and who was later appointed superintendent of the king's new observatory at Richmond; and the apparatus the king commissioned for his own collection from the London maker George Adams in 1761–62.

Other material was added during the collection's long existence. Many of the items are from the seventeenth and eighteenth centuries. Some were owned by the royal family before George III came to the throne, such as the Grand Orrery enlarged for George II. Others were deposited at the Richmond Observatory (now Kew) soon after it was built in 1770. Of particular note are items alleged to have belonged to Robert Boyle.

The observatory building was taken over by the newly founded British Association for the Advancement of Science in the nineteenth century, and a few of its instruments were given to other observatories in Britain. Most of the collection, however, was moved to King's College in London, where it was housed in the King George III Museum opened to the public in 1843. Nineteenth century material was added. Some came from Charles Wheat-

From Public & Private Science

stone, professor of natural philosophy at King's, noted for his work on cable telegraphy. In 1927 the collection was transferred to the Science Museum.

There are about a thousand items in all, described in great detail and profusely photographed. Demainbray's instruments reflect his interest in teaching. There are pulley wheels and model engines (such as Vauloué's pile-driver used in the building of Westminster Bridge in 1738), model pumps, and optical models of short- and long-sightedness, the optics of the camera obscura, and 'Newton's doctrine of light and colour'.

Demainbray's instruments were good, but those of George III were fit for a king. These instruments cover the eighteenth century syllabus of pneumatics and mechanics. The centrepiece is a double-barrelled air pump, based on the design of John Smeaton. The surviving accessories show what caught the public's imagination: brass Magdeburg hemispheres to demonstrate the power exerted by the weight of air; guinea and feather apparatus to demonstrate that they fall at the same rate in a vacuum; and a clockwork bell to demonstrate that sound does not travel in a vacuum.

The king's mechanics collection was also very comprehensive. This subject was made popular by the textbooks of s' Gravesande, Desaguliers and Nollet, which were used by Adams for the design of these pieces. However, this subject could never have been as popular as the spectacular lectures with the air pump, shown so graphically in Joseph Wright's *An Experiment on a Bird in the Air Pump* (exhibited in 1768). This painting was probably inspired by the lectures of James Ferguson. The air pump was the *pièce de résistance* of such scientific shows.

Also fascinating to the public was static electricity, which had only recently come of age. In the 1770s the king sided with his court painter and amateur 'electrician', Benjamin Wilson, against the scientific establishment, deciding that ball-shaped lightning conductors were better than those terminating in sharp points. It is intriguing that there are no electrical instruments in the core of the George III Collection, although there are some fine specimens in the seventeenth and eighteenth century material deposited at Richmond Observatory, including a rare unsigned plate electrical machine from about 1770. This was purchased by the Prince of Wales (later George IV) in the 1780s from George Adams the Younger.

The George III Collection was only the latest in a long tradition of aristocratic collecting in Europe. The study of such material gives an added perspective to our understanding of the scientific activity of the period, augmenting literary sources. It provides specific insights into the manufacture of these devices, the relationship

between instrument-making and scientific practice, and into the changing nature of such collections over time and the science teaching they represent. Such artefacts give us a tangible link with the past, and stimulate the imagination of those who may want to take up science as a career.

It is ironic that in the month of the publication of this lavish catalogue, the Science Museum has announced drastic cuts to its staff, which will lead to the closure of galleries and a loss of experts. In the future there may no longer be the in-house expertise to produce this kind of catalogue. This retrograde step makes the obvious pride expressed in the foreword about restoring the collection "to the prominence it so fully deserves" rather hollow. Perhaps this project will one day be seen as the Science Museum's swan song as it embraces the modern twin notions of interactive science centres and the public understanding of science. Undoubtedly these are worthy means of widening scientific awareness. It should be remembered, however, that historical collections, too, can be important in understanding modern science. But perhaps the present Science Museum management has more faith in the old proverb: 'Happy is the country that has no history'.

Willem Hackmann is at the Museum of the History of Science, Old Ashmolean Building, Oxford OX1 3AZ, UK.

Earth in motion

Paul Segall

Fault Mechanics and Transport Properties of Rocks. Edited by Brian Evans and Teng-Fong Wong. *Academic*: 1992. Pp. 524. £45, \$99.

THIS volume arose out of a 1990 symposium held in honour of William F. Brace on his retirement after three-and-a-half decades at the Massachusetts Institute of Technology. His colleagues and former students have contributed papers that span four subject areas in which Brace made fundamental contributions.

The first section, on brittle failure of crustal rocks, contains papers on experimental studies of faulting. D. A. Lockner and others show how acoustic emissions can be used to map the formation of a new fault surface in a previously unfaulted sample. Other authors discuss fabric development in calcite gouges and the nature of the brittle-ductile transition in feldspar-rich rock. A study of the frictional strength of clay-rich gouges shows that these materials cannot explain the low strength of the San Andreas fault as determined by

the absence of a heat-flow anomaly.

Papers on permeability and flow include G. J. Fischer's new method for computing permeability and storage capacity of rock samples. The method imposes a sinusoidal pressure variation on one end of the sample and relates the amplitude and phase of the pressure signal at the other end to the permeability and storage capacity. Theoretical papers cover permeability in anisotropic rocks and methods for relating hydraulic properties to digitized images of rock cross-sections. Two experimental papers discuss the growth of grain contacts by solution transfer in halite, and the effect of water and carbon dioxide on silicate melt migration through polycrystalline olivine.

The third section, on fracture characterization and physical properties of rock, contains five papers, including one by M. N. Toksöz and colleagues on the *in situ* characterization of fractures from vertical seismic profiling data. The method uses tube waves that are generated when a fracture, which intersects the borehole, is compressed by a passing seismic wave. The dip of the fracture can be related to the efficiency with which tube waves are generated by incident P and SV waves. Two papers in this section report on field observations of joints in siltstones. There is a study of the effect of stress on remanent magnetization, and a study in which X-ray tomography is used to determine the distribution of oil and water in samples during measurements of resistivity.

The final section, on implication of rock mechanics for crustal tectonics, mainly contains theoretical studies, including an already frequently cited paper by J. R. Rice on the weakness of the San Andreas fault. The most obvious explanation for the weakness — high pore-fluid pressures — had previously been dismissed on the grounds that the pressures would hydraulically fracture the adjacent crust. This argument assumes that the state of stress in a mature fault is the same as that in the surrounding crust, but Rice shows how simple mechanical considerations indicate that this is unlikely. Rice's work has re-established high pore-fluid pressure as the most reasonable explanation. Another study in this section argues that creep closure of pores leads to high fluid pressures in the lower crust followed by natural hydrofracturing, resulting in fluid pressure cycling. Other papers address the state of stress on the active faults, and the emplacement of foreland thrust sheets.

The papers are of uniformly high quality and a few represent significant advances. The volume is a fitting tribute to Brace's career.

Paul Segall is in the Department of Geophysics, Stanford University, Stanford, California 94305, USA.

Punctuated equilibrium comes of age

Stephen Jay Gould & Niles Eldredge

The intense controversies that surrounded the youth of punctuated equilibrium have helped it mature to a useful extension of evolutionary theory. As a complement to phyletic gradualism, its most important implications remain the recognition of stasis as a meaningful and predominant pattern within the history of species, and in the recasting of macroevolution as the differential success of certain species (and their descendants) within clades.

PUNCTUATED equilibrium has finally obtained an unambiguous and incontrovertible majority—that is, our theory is now 21 years old¹. We also, with parental pride (and, therefore, potential bias), believe that primary controversy has ceded to general comprehension, and that punctuated equilibrium has been accepted by most of our colleagues (a more conventional form of majority) as a valuable addition to evolutionary theory.

Kellogg¹ began the best book written to celebrate Darwinism's fiftieth birthday by noting how often (and how continually) various critics had proclaimed the *Sterbelager* (death bed) of natural selection. Punctuated equilibrium² has also prospered from announcements of its death or triviality^{3–5} and has been featured in much recent discussion^{6–8}.

As a neonate in 1972, punctuated equilibrium entered the world in unusual guise. We claimed no new discovery, but only a novel interpretation for the oldest and most robust of palaeontological observations: the geologically instantaneous origination and subsequent stability (often for millions of years) of palaeontological 'morphospecies'. This observation had long been ascribed, by Darwin and others, to the notorious imperfection of the fossil record, and was therefore read in a negative light—as missing information about evolution (defined in standard palaeontological textbooks of the time⁹ as continuous anagenetic transformation of populations, or phyletic gradualism).

In a strictly logical sense, this negative explanation worked and preserved gradualism, then falsely equated with evolution itself, amidst an astonishing lack of evidence for this putative main signal of Darwinism. But think of the practical or heuristic dilemma for working palaeontologists: if evolution meant gradualism, and imperfection precluded the observation of such steady change, then scientists could not access the very phenomenon that both motivated their interest and built life's history. As young, committed and ambitious parents, we therefore proposed punctuated equilibrium, hoping to validate our profession's primary data as signal rather than void. We realized that a standard biological account, Mayr's¹⁰ peripatric theory of speciation in small populations peripherally isolated from a parental stock, would yield stasis and punctuation when properly scaled into the vastness of geological time—for small populations speciating away from a central mass in tens or hundreds of thousands of years, would translate in almost every geological circumstance as a punctuation on a bedding plane, not gradual change up a hill of sediment, whereas stasis should characterize the long and recoverable history of successful central populations.

Punctuated equilibrium then grew during its childhood and adolescence, in ways both unruly and orderly. Unruly accidents of history included the misunderstandings of colleagues (who, for example, failed to grasp the key claim about geological scaling, misread geological abruptness as true suddenness, and then interpreted punctuated equilibrium as a saltational theory), and the purposeful misuses of creationist foes as this political issue heated up in the United States during the late 1970s (although we took pride in joining with so many colleagues for a successful fight against this philistine scourge, as one of us testified in the Arkansas 'monkey' trial in 1981¹¹ and the other wrote a book on creationist distortions¹²).

But orderly extensions, implicit in the undeveloped logic of our original argument, fuelled the useful growth of punctuated equilibrium to fruitful adulthood. (We now realize how poorly we initially grasped the implications of our original argument; we thank our colleagues, especially S. M. Stanley¹³ and E. S. Vrba¹⁴, for developing several extensions). We originally focused on tempo, but more important theoretical arguments flowed from implications concerning evolution's mode^{15–17}—particularly the causes surrounding our two major claims for equilibrium, or stasis of established species, and the need to reformulate macroevolution, notably the key phenomenon of trends, as an accumulation of discrete speciation events treated as entities rather than undefinable segments of continua—a subject encompassed by debate about species selection¹³ or species sorting¹⁸.

Punctuated equilibrium and macroevolution

Stasis and its meaning. We opened our original paper with a section on what philosopher N. R. Hanson called "the cloven hoofprint of theory"¹⁹, or the structuring of all supposedly objective observation by expectations of prevailing general views. Stasis, as palpable and observable in virtually all cases (whereas rapid punctuations are usually, but not always, elusive), becomes the major empirical ground for studying punctuated equilibrium. Putting together the philosophical insight of ineluctable theoretical bias, with the empirical theme of the tractability of stasis, we devised a motto: "stasis is data." For no bias can be more constricting than invisibility—and stasis, inevitably read as absence of evolution, had always been treated as a non-subject. How odd, though, to define the most common of all palaeontological phenomena as beyond interest or notice! Yet palaeontologists never wrote papers on the absence of change in lineages before punctuated equilibrium granted the subject some theoretical space. And, even worse, as palaeontologists didn't discuss stasis, most evolutionary biologists assumed continual change as a norm, and didn't even know that stability dominates the fossil record. Mayr has written²⁰: "Of all the claims made in the punctuationalist theory of Eldredge and Gould, the one that encountered the greatest opposition was that of 'pronounced stasis as the usual fate of most species', after having completed the phase of origination... I agree with Gould that the frequency of stasis in fossil species revealed by the recent analysis was unexpected by most evolutionary biologists."

As the most important change in research practice provoked by punctuated equilibrium, stasis has now exited from its closet of non-definition to become a subject of quantitative investigation in all major fossil groups—from microfossils^{21,22} (27,000 measured specimens from 400 closely spaced samples spanning 8 million years in the latter study), to molluscs^{23–27}, to mammals^{28–30}. Although punctuated equilibrium deals directly only with stability of species through time, the higher-level analogue of non-trending in larger clades has also graduated from an undefined non-subject to a phenomenon worth documenting³¹. Moreover, because species often maintain stability through such intense climatic change as glacial cycling³², stasis must be viewed as an active phenomenon, not a passive response to unaltered environments. Many leading evolutionary

theorists, while not accepting our preference for viewing stasis in the context of habitat tracking¹⁷ or developmental constraint^{33,34}, have been persuaded by punctuated equilibrium that maintenance of stability within species must be considered as a major evolutionary problem.

Macroevolution as a problem in species sorting. If punctuated equilibrium has provoked a shift in paradigms for macroevolutionary theory (see ref. 35 for a defence of this view), the main insight for revision holds that all substantial evolutionary change must be reconceived as higher-level sorting based on differential success of certain kinds of stable species, rather than as progressive transformation within lineages (see Eldredge³⁶ on taxic versus transformational views of evolution; Simpson³⁷, however, in the canonical palaeontological statement of the generation before punctuated equilibrium, had attributed 90% of macroevolution to the transformational mode, and only 10% to speciation). Figure 1, our original diagram of punctuated equilibrium, shows how a trend may be produced by differential success of certain species without directional change in any species following its origin.

Darwin's theory of natural selection locates the causality of evolutionary change at one domain on one level: natural selection operating by struggle among individual organisms for reproductive success. Given Darwin's crucial reliance upon lyellian uniformity for extrapolating this mode of change to encompass all magnitudes through all times, the interposition of a level for sorting among stable species breaks this causal reduction and truly, in Stanley's felicitous term³⁸, "decouples" macro- from microevolution. Decoupling is not a claim for the falseness or irrelevancy of microevolutionary mechanisms, especially natural selection, but a recognition that darwinian extrapolation cannot fully explain large-scale change in the history of life.

The main point may be summarized as follows. Most macroevolution must be rendered by asking what kinds of species within a clade did better than others (speciated more frequently, survived longer), or what biases in direction of speciation prevailed among species within a clade. Such questions enjoin a very different programme of research from the traditional 'how did natural selection within a lineage build substantial adaptation during long stretches of time?' The new questions require a direct study of species and their differential success; older queries focused downward upon processes within populations and their extrapolation through time. Darwin's location of causality in organisms must be superseded by a hierarchical model of selection, with simultaneous and important action at genic, organismal and taxic levels^{39,40}. Williams³⁴, who so stoutly defended classical Darwinism against older, invalid, and very different forms of group selection⁴¹, now acknowledges the importance of such clade selection in macroevolution. Punctuated equilibrium has been used as a central concept in the development of hierarchy theory in evolutionary biology.

Implications. Any theory with a claim to novelty in broad perspective must enlighten old problems and suggest extensions. The speciation view of macroevolution, which does not strictly require punctuated equilibrium, but which was nurtured and has thrived in its context, requires a reformulation of nearly all macroevolutionary questions. For example, so-called living fossils, once treated as lineages rendered static by optimal adaptation, unusually stable environment, or lack of genetic variation, should be reconceptualized as members of groups with unusually low speciation rates, and therefore little opportunity to accumulate change⁴². (We have no evidence that the species of 'living fossil' groups are particularly old. For example, the western Atlantic horseshoe crab, *Limulus polyphemus*—the 'type example' of the phenomenon—has no fossil record at all, whereas the genus can only be traced to the Miocene.)

Going further, the entire tradition of expressing evolutionary change in darwin units (where 1 darwin equals character change by a factor of e in 1 million years)⁴³ makes no sense in a speciation context. (If a lineage goes from species A to D in 10

million years through three episodes of rapid change with intervening stasis, a cited rate of so many millidarwins becomes a meaningless average.) We learn as a received truth of evolution, for example, that human brain size increased at an extraordinary (many say unprecedented) rate during later stages of our lineage. But this entrenched belief may be a chimaera born of an error in averaging rates over both punctuations and subsequent periods of stasis. *Homo sapiens* is a young species, perhaps no more than 200,000 years old. If most of our increment accrued quickly at our origin, but we then express this entirety from our origin to the present time as a darwin rate, we calculate a high value because our subsequent time of stasis has been so short. But if the same speciation event, with the same increment in the same time, had occurred two million years ago (with subsequent stasis), the darwin rate for the identical event would be much lower.

Cope's rule, the tendency for phyletic increase in body size, had generally been attributed to selective value of large size within anagenetic lineages, but is probably better interpreted^{44,45} as greater propensity for speciation in smaller species, for whom increasing size is the only 'open' pathway (see Martin⁴⁶ on the negative correlation of generic species richness and body size). Raup and Sepkoski⁴⁷ proposed a conventional explanation for decreasing rate of background extinction through geological time: generally better adaptation of later species. But Valentine⁴⁸ and Gilinsky (personal communication) offer an interesting speciation alternative: if extinction intensities were constant through time, groups with equally high speciation and extinction

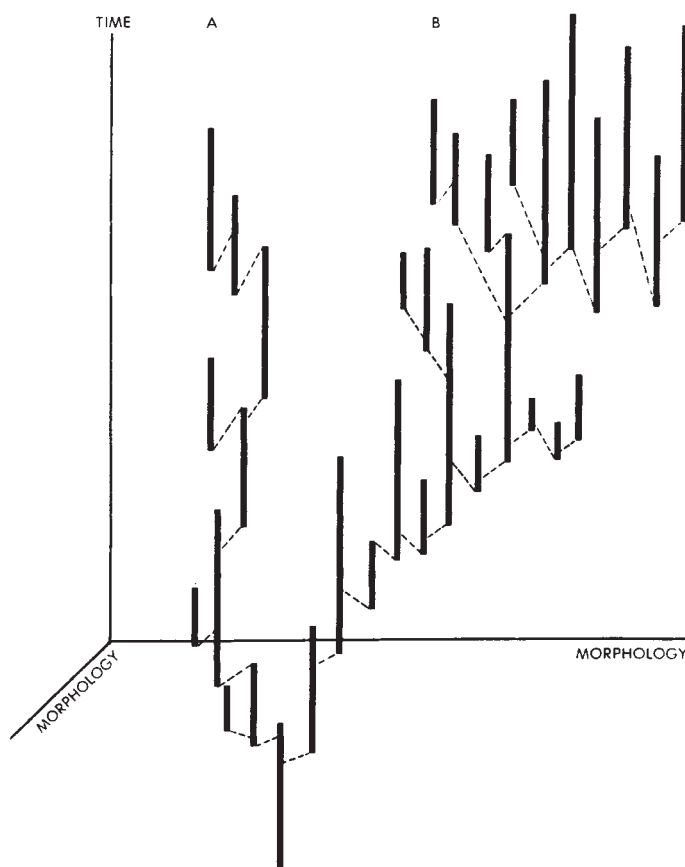


FIG. 1 Three-dimensional sketch contrasting a pattern of relative stability (A) with a trend (B), where speciation (dashed lines) is occurring in both major lineages. Morphological change is depicted here along the horizontal axes, while the vertical axis is time. Though a retrospective pattern of directional selection might be fitted as a straight line in (B), the actual pattern is stasis within species, and differential success of species exhibiting morphological change in a particular direction. For further explanation, see ref. 1.

rates would fare just as well as groups with equally lower rates. But extinction intensities vary greatly, and the geological record features episodes of high dying, during which extinction-prone groups are more likely to disappear, leaving extinction-resistant groups as life's legacy. Valentine concludes that "these clade-characteristic rates are of course not adaptations *per se*, but effects flowing from clade properties that were established probably during the early radiations that founded the clades."

The most exciting direct extensions of punctuated equilibrium now involve the study of correlated punctuational events across taxa, and the ecological and environmental sources of such cohesion. Eminently testable are Vrba's⁴⁹ "turnover-pulse hypothesis" of evolution concentrated in punctuational bursts at times of worldwide climatic pulsing, one of which, about 2.5 million years ago, may have stimulated the origin of the genus *Homo*; and Brett's⁵⁰ hypothesis of "coordinated stasis" for the structuring of major palaeontological faunas. What might be the ecological source of such striking coherence across disparate taxa through such long times?⁵¹

The empirics of punctuated equilibrium

Like all major theories in the sciences of natural history, including natural selection itself, punctuated equilibrium is a claim about relative frequency, not exclusivity⁵². Phyletic gradualism has been well documented, again across all taxa from microfossils⁵³ to mammals^{54,55}. Punctuated equilibrium surely exists in abundance, but validation of the general hypothesis requires a relative frequency sufficiently high to impart the predominant motif and signal to life's history. The issue remains unsettled, but we consider (in our biased way) that four classes of evidence establish a strong putative case for punctuated equilibrium in this general sense.

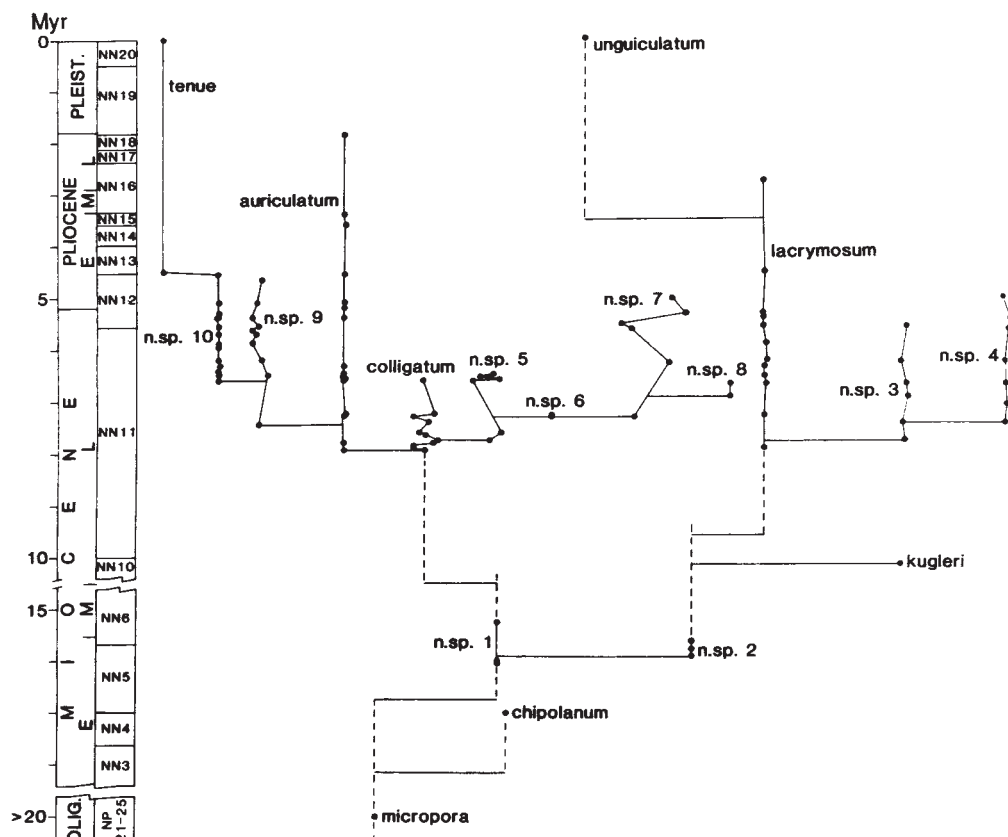
■ **Individual cases.** Examples of stasis alone (cited earlier) and simple abrupt replacement, although conforming to expectations of punctuated equilibrium, are not direct evidence for our mechanism: for stasis might just be a lull in anagenetic gradual-

ism (though pervasive stasis for long periods in all species of a fauna (a common finding) would require special pleading from gradualists), and replacement might represent rapid transformation without branching, or migration of a distant (phyletic or geographic) relative rather than evolution *in situ*. A good test of punctuated equilibrium requires (in addition to the obvious need for documented rapidity in an interval known to be sufficiently short) both a phyletic hypothesis to assert sister-group relationship of the taxa involved, and survival of putative ancestors to affirm an event of true branching rather than rapid phyletic transformation.

Given these stringent requirements, and in the light of such an imperfect fossil record, we are delighted that so many cases have been well documented, particularly in the crucial requirement of ancestral survival after punctuated branching^{32,56-61}. Williamson's discovery⁶² of multiple molluscan speciation events in isolated African lakes, with return of ancestral lineages upon reconnection with parental water bodies, has been most widely discussed⁶³ and disputed⁶⁴ (although all accept the punctuational pattern). Cheetham's elegant and meticulously documented^{65,66} story of evolution in the bryozoan *Metrarabdotos* from Tertiary strata of the Caribbean is particularly gratifying (Fig. 2) in the number of purely punctuational events, the full coverage of the lineage, and the unusual completeness of documentation, especially as Cheetham began his study expecting to reconfirm a gradualistic interpretation (writing to McKinney: "The chronocline I thought was represented . . . is perhaps the most conspicuous casualty of the restudy, which shows that the supposed cline members largely overlap each other in time. Eldredge and Gould were certainly right about the danger of stringing a series of chronologically isolated populations together with a gradualist's expectations.")

On the subject of punctuational corrections for received gradualistic wisdom, Prothero and Shubin⁶⁷ have shown that the most 'firmly' gradualistic part of the horse lineage (the general, and false, exemplar of gradualism in its totality), the Oligocene

FIG. 2 Reproduced with permission from ref. 65. Phylogenetic (stratophenetic) tree of Caribbean species of *Metrarabdotos*. Relationships were constructed by calculating euclidean distances between sampled populations using all canonical scores from discriminant analysis, and connecting nearest morphological neighbours in stratigraphic sequence. Morphological distances (horizontal axis) between inferred ancestor and descendant species are to scale; those between species of different pairs are not necessarily so (for example, *M. kugleri* is morphologically more distant from *M. n. sp. 3* than from its ancestor, *M. n. sp. 2*, even though the diagram makes it appear otherwise). Within-species fluctuation in species-distinguishing morphology is based on pairwise discriminant scores scaled to the distances between species pairs.



transition from *Mesohippus* to *Miohippus*⁶⁸, conforms to punctuated equilibrium, with stasis in all species of both lines, transition by rapid branching rather than phyletic transformation, and stratigraphic overlap of both genera (one set of beds in Wyoming has yielded three species of *Mesohippus* and two of *Miohippus*, all contemporaries). Prothero and Shubin conclude: "This is contrary to the widely-held myth about horse species as gradually-varying parts of a continuum, with no real distinctions between species. Throughout the history of horses, the species are well-marked and static over millions of years. At high resolution, the gradualistic picture of horse evolution becomes a complex bush of overlapping, closely related species."

■ **Relative frequencies.** Elegant cases don't make punctuated equilibrium any more than a swallow makes a summer, but there are a growing number of reports documenting an overwhelming relative frequency (often an exclusivity) for punctuated equilibrium in entire groups or faunas. Consider the lifetime testimonies of taxonomic experts on microfossils⁶⁹, on brachiopods^{70,71} and on beetles⁷². Fortey⁷³ has concluded for trilobites and graptolites "that the gradualistic mode does occur especially in pelagic or planktic forms, but accounts for 10% or less of observations of phyletic change, and is relatively slow".

Other studies access all available lineages in entire faunas and assert the dominance of punctuated equilibrium. Stanley and Yang²⁶ found no gradualism at all in the classic Tertiary molluscan sequences of the Gulf and Atlantic Coasts. With the exception of *Gryphaea*, Hallam²³ detected no phyletic change in shape (but only for body size) in any Jurassic bivalve in Europe. Kelley^{24,25} documented the prevalence of punctuation for molluscs in the famous Maryland Miocene sequence, and Vrba⁷⁴ has done the same for African bovids. Even compilations from the literature, so strongly biased by previous traditions for ignoring stasis as non-data and only documenting putative gradualism, grant a majority to punctuated equilibrium, as in Barnovsky's⁷⁵ compendium for Quaternary mammals, with punctuated equilibrium "supported twice as often as phyletic gradualism . . . the majority of species considered exhibit most of their morphological change near a speciation event, and most species seem to be discrete entities". When controlled studies are done by one team in the field, punctuated equilibrium almost always seems to predominate. Prothero⁷⁶ "examined all the mammals with a reasonably complete record from the Eocene-Oligocene beds of the Big Badlands of South Dakota and related areas in Wyoming and Nebraska . . . With one exception (gradual dwarfing in the oreodont *Miniochoerus*), we found that all of the Badlands mammals were static through millions of years, or speciated abruptly (if they changed at all)."

■ **Inductive patterns.** Even more general inductive patterns should be explored as criteria. Stanley^{38,77} has proposed a series of tests, all carried out to punctuated equilibrium's advantage. Others suggest that certain environments and ecologies should be conducive to one preferred mode along the continuum of possibilities. Johnson^{78,79} suggests that punctuated equilibrium should dominate in the benthic environments that yield most of our fossil record, while gradualism might prevail in pelagic realms. Sheldon⁸⁰ proposes the counter-intuitive but not unreasonable idea that punctuated equilibrium may prevail in unstable environments, gradualism in stable regimes.

■ **Tests from living organisms.** Distinct evolutionary modes yield disparate patterns as results; punctuated equilibrium might therefore be tested by studying the morphological and taxonomic distributions of organisms, including living faunas. (Several of Stanley's tests³⁸ use modern organisms, and other criteria from fossils should be explored—especially the biometric discordance or orthogonality, favourable to punctuated equilibrium and actually found where investigated^{25,81}, of within and between species trends.)

Cladistic patterns should provide a good proving ground. Avise⁸² performed an interesting and much quoted test, favourable to gradualism, by comparing genetic and morphological

differences in two fish clades of apparently equal age and markedly different speciation frequencies. But as Mayden⁸³ argued, this test was wrong in its particular case, and non-optimal as a general procedure; a better method would compare cladistic sister groups, guaranteed by this status to be equal in age. Mindel *et al.*^{84,85} have now performed such a test on the reptilian genus *Sceloporus* and on allozymic data in general, and have validated punctuated equilibrium's key claim for positive correlation of evolutionary distance and speciation frequency. Lemen and Freeman's⁸⁶ interesting proposal for additional cladistic tests cannot be sustained because they must assume that unbranched arms of their cladograms truly feature no speciation events along their routes, whereas numerous transient and extinct species must populate most of these pathways. Wagner⁸⁷ has developed a way of estimating rapidly branching speciation versus gradual speciation or transformation from cladograms, and his initial results favour predominant rapid branching in Palaeozoic gastropods.

Difficulties and prospects

Many semantic and terminological muddles that once impeded resolution of this debate have been clarified. Opponents now accept that punctuated equilibrium was never meant as a saltational theory, and that stasis does not signify rock-hard immobility, but fluctuation of little or no accumulated consequence, and temporal spread within the range of geographic variability among contemporary populations—by Stanley's proper criterion, so strikingly validated in his classic study²⁶. We trust that everyone now grasps the centrality of relative frequency as a key criterion (and will allow, we may hope, that enough evidence has now accumulated to make a case, if not fully prove the point).

Evolutionary biologists have also raised a number of theoretical issues from their domain of microevolution. Some, like the frequency of sibling speciation, seem to us either irrelevant or untroublesome as a bias against, rather than for, our view (as we then underestimate the amount of true speciation from palaeontologically defined morphospecies, and such an underestimate works against punctuated equilibrium). Others, like the potential lack of correspondence between biospecies and palaeontological morphospecies, might be worrisome, but available studies, done to assess the problem in the light of punctuated equilibrium, affirm the identity of palaeontological taxa with true biospecies (see Jackson and Cheetham⁸⁸ on bryozoan species, and Michaux²⁷ on palaeontological stasis in gastropod morphospecies that persist as good genetic biospecies).

But continuing unhappiness, justified this time, focuses upon claims that speciation causes significant morphological change, for no validation of such a position has emerged (while the frequency and efficacy of our original supporting notion, Mayr's "genetic revolution" in peripheral isolates, has been questioned). Moreover, reasonable arguments for potential change throughout the history of lineages have been advanced^{6,34}, although the empirics of stasis throws the efficacy of such processes into doubt. The pattern of punctuated equilibrium exists (at predominant relative frequency, we would argue) and is robust. *Eppur non si muove*; but why then? For the association of morphological change with speciation remains as a major pattern in the fossil record.

We believe that the solution to this dilemma may be provided in a brilliant but neglected suggestion of Futuyma⁸⁹. He holds that morphological change may accumulate anywhere along the geological trajectory of a species. But unless that change be "locked up" by acquisition of reproductive isolation (that is, speciation), it cannot persist or accumulate and must be washed out during the complexity of interdigitation through time among varying populations of a species. Thus, species are not special because their origin permits a unique moment for instigating change, but because they provide the only mechanism for protecting change. Futuyma writes: "In the absence of reproductive

isolation, differentiation is broken down by recombination. Given reproductive isolation, however, a species can retain its distinctive complex of characters as its spatial distribution changes along with that of its habitat or niche . . . Although speciation does not accelerate evolution within populations, it provides morphological changes with enough permanence to be registered in the fossil record. Thus, it is plausible to expect many evolutionary changes in the fossil record to be associated with speciation." By an extension of the same argument, sequences of speciation are then required for trends: "Each step has had a more than ephemeral existence only because reproductive isolation prevented the slippage consequent on interbreeding with other populations . . . Speciation may facilitate anagenesis by retaining, stepwise, the advances made in any one direction." Futuyma's simple yet profound insight may help to heal the remaining rifts and integrate punctuated equilibrium into an evolutionary theory hierarchically enriched in its light^{17,18}.

In summarizing the impact of recent theories upon human concepts of nature's order, we cannot yet know whether we have witnessed a mighty gain in insight about the natural world (against anthropocentric hopes and biases that always hold us

down), or just another transient blip in the history of correspondence between misperceptions of nature and prevailing social realities of war and uncertainty. Nonetheless, contemporary science has massively substituted notions of indeterminacy, historical contingency, chaos and punctuation for previous convictions about gradual, progressive, predictable determinism. These transitions have occurred in field after field; Kuhn's⁹⁰ celebrated notion of scientific revolutions is, for example, a punctuational theory for the history of scientific ideas. Punctuated equilibrium, in this light, is only palaeontology's contribution to a *Zeitgeist*, and *Zeitgeists*, as (literally) transient ghosts of time, should never be trusted. Thus, in developing punctuated equilibrium, we have either been toadies and panderers to fashion, and therefore destined for history's ashheap, or we had a spark of insight about nature's constitution. Only the punctuational and unpredictable future can tell. □

Stephen Jay Gould is at the Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA; and Niles Eldredge is at the Department of Invertebrates, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024, USA.

- Kellogg, V. L. *Darwinism Today* (Bell, London, 1907).
- Eldredge, N. & Gould, S. J. in *Models in Paleobiology* (ed. Schopf, T.J.M.) 82–115 (Freeman, Cooper, San Francisco, 1972).
- Dawkins, R. *The Blind Watchmaker* (Norton, New York, 1986).
- Levinton, J. *Genetics, Paleontology and Macroevolution* (Cambridge Univ. Press, New York, 1988).
- Hoffman, A. *Arguments on Evolution* (Oxford Univ. Press, New York, 1989).
- Ridley, M. *Evolution* (Blackwell Scientific, Boston, 1993).
- Erwin, D. Speciation in the Fossil Record 138–140 (Geol. Soc. Am. Ann. Meeting, Abstr., Boulder, Colorado, 1992).
- Mayr, E., Boulding, K. E. & Masters, R. D. in *The Dynamics of Evolution: The Punctuated Equilibrium Debate in the Natural and Social Sciences* (eds Somit, A. & Peterson, S. A.) (Cornell Univ. Press, Ithaca, NY, 1992).
- Moore, R. C., Lalicker, C. G. & Fischer, A. G. *Invertebrate Fossils* (McGraw-Hill, New York, 1952).
- Mayr, E. *Animal Species and Evolution* (Harvard Univ. Press, Cambridge, MA, 1963).
- Gould, S. J. *Heri's Teeth and Horse's Toes* (Norton, New York, 1983).
- Eldredge, N. *The Monkey Business. A Scientist Looks at Creationism* (Pocket Books, New York, 1982).
- Stanley, S. M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 646–650 (1975).
- Vrba, E. S. *Afr. J. Sci.* **7**, 61–84 (1980).
- Gould, S. J. & Eldredge, N. *Paleobiology* **3**, 115–151 (1977).
- Gould, S. J. in *Perspectives on Evolution* (ed. Milkman, R.) 83–104 (Sinauer, Sunderland, MA, 1982).
- Eldredge, N. *Macroevolutionary Dynamics* (McGraw-Hill, New York, 1989).
- Vrba, E. & Gould, S. J. *Paleobiology* **12**, 217–228 (1986).
- Hanson, N. R. *Perception and Discovery* (Freeman, Cooper, San Francisco, 1969).
- Mayr, E. in *The Dynamics of Evolution* (eds Somit, A. & Peterson, S. A.) 21–53 (Cornell Univ. Press, Ithaca, NY, 1992).
- Sorhannus, U. *Hist. Biol.* **31**, 241–247 (1990).
- Thomas, E. *Utrecht Micropal. Bull.* **23**, 1–167 (1980).
- Hallam, A. *Paleobiology* **4**, 16–25 (1978).
- Kelley, P. H. J. *Paleontol.* **57**, 581–598 (1983).
- Kelley, P. H. J. *Paleontol.* **58**, 1235–1250 (1984).
- Stanley, S. M. & Yang, X. *Paleobiology* **13**, 113–139 (1987).
- Michaux, B. *Biol. J. Linn. Soc.* **38**, 239–255 (1989).
- West, R. M. *Paleobiology* **5**, 252–260 (1979).
- Schankler, D. M. *Nature* **293**, 135–138 (1981).
- Lich, D. K. *Paleobiology* **16**, 384–395 (1990).
- Budd, A. F. & Coates, A. G. *Paleobiology* **18**, 425–446 (1992).
- Cronin, T. M. *Science* **227**, 60–63 (1985).
- Maynard Smith, J. A. *Rev. Genet.* **17**, 11–25 (1984).
- Williams, G. C. *Natural Selection: Domains, Levels and Challenges* (Oxford Univ. Press, New York, 1992).
- Ruse, M. in *The Dynamics of Evolution* (eds Somit, A. & Peterson, S. A.) 139–168 (Cornell Univ. Press, Ithaca, NY, 1992).
- Eldredge, N. *Bull. Carnegie Mus. nat. Hist.* **13**, 7–19 (1979).
- Simpson, G. G. *The Major Features of Evolution* (Columbia Univ. Press, New York, 1953).
- Stanley, S. M. *Macroevolution* (Freeman, San Francisco, 1979).
- Lloyd, E. A. *The Structure and Confirmation of Evolutionary Theory* (Greenwood, New York, 1988).
- Brandon, R. N. *Adaptation and Environment* (Princeton Univ. Press, Princeton, NJ, 1990).
- Williams, G. C. *Adaptation and Natural Selection* (Princeton Univ. Press, Princeton, NJ, 1966).
- Eldredge, N. & Stanley, S. M. *Living Fossils* (Springer, New York, 1984).
- Haldane, J. B. S. *Evolution* **3**, 51–56 (1949).
- Stanley, S. M. *Evolution* **27**, 1–26 (1973).
- Gould, S. J. *J. Paleont.* **62**, 319–329 (1988).
- Martin, R. A. *Hist. Biol.* **6**, 73–90 (1992).
- Raup, D. M. & Sepkoski, J. J. Jr *Science* **215**, 1501–1503 (1982).
- Valentine, J. W. in *Molds, Molecules and Metazoa* (eds Grant, P. R. & Horn, H. S.) 17–32 (Princeton Univ. Press, Princeton, NJ, 1992).
- Vrba, E. S. *Suid-Afrikaanse Tydskrif Wetens.* **81**, 229–236 (1985).
- Brett, C. E. *Coordinated Stasis and Evolutionary Ecology of Silurian–Devonian Marine Biotas in the Appalachian Basin* 139 (Geol. Soc. Am. Ann. Meeting, Abstr., Boulder, Colorado, 1992).
- Morris, P. J., Ivany, L. & Schopf, K. M. *Paleoecological Stasis in Evolutionary Theory* 313 (Geol. Soc. Am. Ann. Meeting, Boulder, Colorado, 1992).
- Gould, S. J. *Am. Sci.* **74**, 60–69 (1986).
- MacLeod, N. *Paleobiology* **17**, 167–188 (1999).
- Gingerich, P. D. *Am. J. Sci.* **276**, 1–28 (1976).
- Chaline, J. & Laurin, B. *Paleobiology* **12**, 203–216 (1986).
- Bergström, J. & Levi-Setti, R. *Geol. Palaeontol.* **12**, 1–40 (1978).
- Smith, A. B. & Paul, C. R. C. *Sp. Pap. Palaeont.* **33**, 29–37 (1985).
- Flynn, L. J. *Contr. Geol. Univ. Wyoming* **3**, 273–285 (1986).
- Finney, S. C. *Geol. Soc. Sp. Pub.* **20**, 103–113 (1986).
- Wei, K. Y. & Kennett, J. P. *Paleobiology* **14**, 345–363 (1988).
- Nehm, R. H. & Geary, D. H. *Paleontol. Soc. Sp. Pub.* **6**, 222 (1992).
- Williamson, P. G. *Nature* **293**, 437–443 (1981).
- Williamson, P. G. *Biol. J. Linn. Soc.* **28**, 307–324 (1985).
- Fryer, G., Greenwood, P. H. & Peake, J. F. *Biol. J. Linn. Soc.* **20**, 195–205 (1983).
- Cheetham, A. H. *Paleobiology* **12**, 190–202 (1986).
- Cheetham, A. H. *Paleobiology* **13**, 286–296 (1987).
- Prothero, D. R. & Shubin, N. in *The Evolution of Perissodactyls* (eds Prothero, D. R. & Schoch, R. M.) 142–175 (Oxford Univ. Press, Oxford, 1989).
- Simpson, G. G. *Horses* (Oxford Univ. Press, Oxford, 1951).
- MacGillivray, H. J. *Bijdragen tot de Dierkunde* **38**, 69–74 (1968).
- Ager, D. V. *Proc. Geologists' Ass.* **87**, 131–160 (1976).
- Ager, D. V. *Palaeontology* **26**, 555–565 (1983).
- Coope, G. R. A. *Rev. Ecol. Syst.* **10**, 247–267 (1979).
- Fortey, R. A. *Sp. Pap. Palaeontol.* **33**, 17–28 (1985).
- Vrba, E. S. in *Living Fossils* (eds Eldredge, N. & Stanley, S. M.) 62–79 (Springer, New York, 1984).
- Barnovsky, A. D. *Curr. Mammal.* **1**, 109–147 (1987).
- Prothero, D. R. *Skeptic* **1**, 38–47 (1992).
- Stanley, S. M. *Paleobiology* **4**, 26–40 (1978).
- Johnson, J. G. *Paleontol.* **49**, 646–661 (1975).
- Johnson, J. G. *Paleontol.* **56**, 1329–1331 (1982).
- Sheldon, P. R. *Nature* **345**, 772 (1990).
- Shapiro, E. A. *Natural Selection in a Miocene Pectinid: A Test of the Punctuated Equilibria Model* 490 (Geol. Soc. Am. Ann. Meeting, Abstr., Boulder, Colorado, 1978).
- Avise, J. C. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5083–5087 (1977).
- Mayden, R. L. *Syst. Zool.* **35**, 591–602 (1986).
- Mindell, D. P., Sites, J. R. Jr & Graur, D. *Cladistics* **5**, 49–61 (1989).
- Mindell, D. P., Sites, J. W. Jr & Graur, D. J. *evol. Biol.* **3**, 125–131 (1990).
- Lemen, C. A. & Freeman, P. W. *Evolution* **43**, 1538–1554 (1989).
- Wagner, P. J. *Cladograms as Tests of Speciation Patterns* 139 (Geol. Soc. Am. Ann. Meeting, Abstr., Boulder, Colorado, 1992).
- Jackson, S. B. C. & Cheetham, A. H. *Science* **248**, 579–583 (1990).
- Futuyma, D. J. *Am. Nat.* **130**, 465–473 (1987).
- Kuhn, T. S. *The Structure of Scientific Revolutions* (Univ. Chicago Press, Chicago, 1962).

The reaction cycle of GroEL and GroES in chaperonin-assisted protein folding

Jörg Martin, Mark Mayhew, Thomas Langer* & F. Ulrich Hartl†

Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA

The reaction mechanism of protein folding by the chaperonin GroEL and its regulator GroES has been defined. GroES and substrate protein counteract each other's effects on GroEL: whereas GroES stabilizes GroEL in the ADP-bound state, binding of unfolded polypeptide within the cavity of the GroEL cylinder triggers ADP and GroES release. Upon ADP-ATP exchange, GroES reassociates with GroEL and ATP hydrolysis discharges the bound protein for folding. Partially folded protein rebinds to the chaperonin, thus perpetuating the cycle until folding is complete.

MOLECULAR chaperones comprise a number of proteins that share the general property of interacting with other proteins in their non-native conformations (for recent reviews, see refs 1–3). Their basic function appears to be the prevention of incorrect associations within and between polypeptide chains during *de novo* protein folding and the protection of pre-existing proteins under cellular stress. Certain chaperones mediate efficient protein folding by releasing and rebinding their substrate in a controlled, ATP-dependent reaction. Representative of this class of chaperones are the members of the heat-shock protein Hsp60 family. These so-called chaperonins⁴ are large oligomeric structures composed of two stacked heptameric rings of ~60K subunits that form a central cavity^{5,6}. The *Escherichia coli* chaperonin GroEL binds 1–2 molecules of unfolded protein in this cavity^{7,8}, presumably in the conformation of a compact folding intermediate^{9,10}. Formation of native tertiary structure occurs upon ATP-dependent discharge of the bound protein. The 14 ATPase sites of GroEL act in a coordinated manner in this process^{11–13}, which is dependent on the regulatory function of the co-chaperonin GroES^{9,14–16}. GroES is a single heptameric ring of subunits of M_r 10K (refs 4, 17), and it forms a 1:1 complex with GroEL by binding to one end of the GroEL cylinder^{8,18,19}. This inhibits the GroEL ATPase^{9,17,20}, whereas the association of unfolded protein markedly stimulates the hydrolysis of ATP by GroEL^{9,13}.

During folding, the polypeptide substrate apparently remains within the GroEL cylinder and is thus protected from unproductive interactions with other folding chains⁹. This important property of GroEL is dependent on GroES. With proteins such as rhodanese, whose spontaneous renaturation is inefficient, ATP-dependent release from GroEL in the absence of GroES results in aggregation rather than productive folding^{9,21}. How GroES exerts its regulatory function is key to understanding the mechanism of chaperonin action. Focusing on the interaction between GroEL and GroES, we demonstrate the intriguing principle of the reaction: GroES and folding protein oppose each other's effects on GroEL. Both GroES and substrate protein go through ATP-dependent cycles of transient release from GroEL. These cycles are mechanistically coupled so that the reaction is self-perpetuating until protein folding is complete. In the accompanying letter we show that GroES itself binds ATP, which may be involved in regulating the ATPase function of GroEL²².

Cycles of GroES release and rebinding

Complex formation between GroEL and GroES occurs in the

presence of Mg-ATP or Mg-ADP, as reflected by the cofractionation of the ~800K GroEL and the ~70K GroES on a size-exclusion column (Fig. 1a). Given that unfolded protein and GroES have opposite effects on the GroEL ATPase, we tested whether the GroEL-GroES interaction remains stable during ATP-dependent protein folding. Unfolded rhodanese (M_r 33K) or citrate synthase (M_r 50K), which both require GroES for their refolding by GroEL^{9,15,21}, were diluted from 6 M guanidinium-Cl into buffer solution containing the GroEL-GroES complex and nucleotide. The reactions were immediately analysed by size-exclusion chromatography. In the case of rhodanese addition, GroES eluted as a complex with GroEL. This was observed with either ADP or ATP in the column buffer. In the presence of Mg-ATP, active rhodanese was recovered, fractionating over a broad range from the expected position of the native protein to the GroEL peak (fractions 8 and 3, respectively; not shown). This rhodanese had been released from GroEL and refolded during column fractionation while the GroEL-GroES complex apparently remained intact.

In contrast, when unfolded citrate synthase was added to GroEL-GroES, about 30–40% of GroES was displaced from GroEL in the presence of Mg-ADP or Mg-ATP (Fig. 1a). The separation of GroES from GroEL precluded the correct folding and dimer assembly of citrate synthase. No active enzyme was eluted from the column, although more than 50% of the citrate synthase had been released from GroEL. This released protein was only recovered by a column wash with 8 M urea (not shown). On the other hand, when GroEL, GroES and unfolded citrate synthase were first incubated with Mg-ATP until folding/assembly was complete, all of the GroES co-eluted with GroEL on the column (Fig. 1a). Upon binding to GroEL-GroES in the presence of Mg-ADP, both unfolded rhodanese and citrate synthase were preserved in a conformation fully competent for subsequent ATP-dependent reactivation (Fig. 1b).

These results indicate that the association of unfolded substrate protein decreases the affinity of GroEL for GroES. Accordingly, GroES binding should have a similar effect on the interaction between GroEL and substrate protein. This could be confirmed when GroES was added in the presence of ADP to a complex of GroEL and unfolded dihydrofolate reductase, a protein with relatively low affinity for GroEL (J.M. and F.U.H., manuscript in preparation).

We further addressed the possibility that GroES cycles between bound and free states during the folding reaction. Once released, GroES would be free to interact with any GroEL molecule in the solution. This was tested in the following experiment (Fig. 2): an excess of denatured rhodanese (P_{1U}) was added to GroEL. As expected, addition of Mg-ATP and of GroES (equimolar with GroEL) resulted in efficient reactivation

* Present address: Institut für Physiologische Chemie, Goethestrasse 33, 80336 München, Germany.

† To whom correspondence should be addressed.

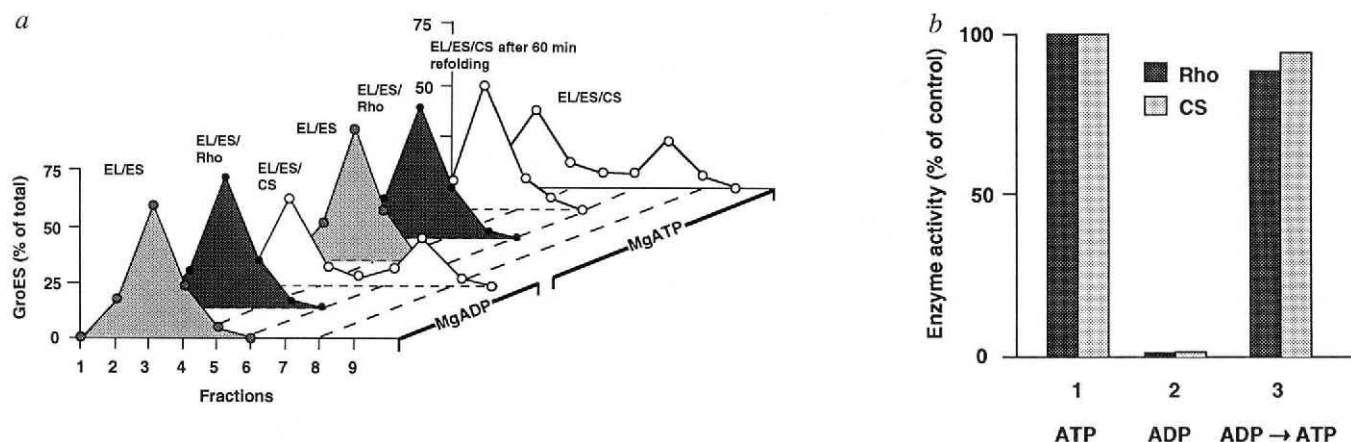


FIG. 1 a, Size fractionation of GroES in the presence of GroEL, nucleotides and unfolded protein as indicated. b, Reactivation of denatured rhodanese (Rho) and citrate synthase (CS) upon binding to a GroEL–GroES–Mg-ADP complex and subsequent addition of ATP. Refolding obtained by adding denatured protein to GroEL–GroES in the presence of Mg-ATP is set to 100%.

METHODS. GroEL and GroES proteins were purified from an overproducing strain of *E. coli* bearing the plasmid pOF39 (refs 9, 29). In a, GroEL 14-mer (0.15 μ M) and 3 H-labelled⁸ GroES 7-mer (0.12 μ M) were incubated (10 min at 23 °C) for complex formation in buffer A (20 mM Tris, pH 7.4, 100 mM KCl) containing 5 mM magnesium acetate. ATP or ADP were present at 1 mM where indicated. Each incubation was divided into three reactions: reaction 1 was analysed without further addition (EL/ES); reaction 2 received denatured bovine rhodanese (1.2 μ M) (EL/ES/Rho), and reaction 3 denatured porcine citrate synthase (1.2 μ M) (EL/ES/CS) (final concentrations). Rho and CS were unfolded in 6 M guanidinium-Cl in buffer A with 3 mM dithiothreitol for 1 h at 23 °C and were 100-fold diluted into GroEL/GroES-containing solution by rapid

manual mixing. Sodium thiosulphate (7.5 mM) and oxaloacetic acid (0.5 mM) were present in all Rho and CS reactions, respectively, to stabilize refolded protein. After 30 s, reactions were applied onto a 2.5-ml Sephacryl-S300 gel filtration column (Pharmacia) equilibrated in buffer A containing the corresponding nucleotide. An additional reaction containing GroEL, GroES, denatured CS and ATP was incubated for 60 min before separation on the size exclusion column. Column fractions were analysed by SDS-PAGE, fluorography and laser densitometry⁹. When GroEL and GroES were applied on their own, they peaked in fractions 3 and 7/8, respectively (not shown). In b, denatured Rho or CS was diluted into buffer A containing GroEL (0.45 μ M), GroES (0.45 μ M), magnesium acetate (5 mM), and either 0.2 mM ATP (reaction 1) or 0.2 mM ADP (reactions 2 and 3). After 10 min, reactions 1 and 3 received 2 mM ATP. Enzyme activities were determined after 50 min incubation at 23 °C (refs 9, 21, 30). The 100% activities correspond to 70% and 65% of the total Rho and CS present in the reactions, respectively, compared with equivalent amounts of the native enzymes.

(lane 1), whereas addition of nucleotide alone did not (lane 2). When the GroES was supplied as a preformed GroEL–GroES complex, only inefficient folding was observed, indicating that both rhodanese and GroES remained largely associated with their respective GroEL (lane 3). But when GroES was added together with GroEL and unfolded citrate synthase (P_{2U}), 60% of the maximal rhodanese reactivation was recorded (lane 4). As already observed, GroES was released from GroEL–citrate synthase and was able to interact with the GroEL–rhodanese complex, driving the folding of rhodanese.

Assuming a common principle of chaperonin action, release and rebinding of GroES should also occur with rhodanese as the substrate, although rhodanese did not cause the separation of GroES from GroEL upon column fractionation (Fig. 1). Therefore, in the converse experiment, unfolded citrate synthase (P_{1U}) was bound to GroEL and the effects of adding preformed complexes of GroEL and GroES or of GroEL, GroES and rhodanese (P_{2U}) were tested (Fig. 2). Only in the latter case was efficient folding of citrate synthase observed. Thus, unfolded rhodanese also induces the transient but complete release of GroES from GroEL. Citrate synthase apparently destabilizes the GroEL–GroES complex more than rhodanese does, but with both proteins GroES cycles between bound and free states during GroEL-mediated folding. Consistent with the number of ATP molecules hydrolysed in the reaction⁹, a single round of GroEL–GroES interaction is not sufficient for these proteins to fold.

Asymmetrical binding of nucleotide and GroES

Binding of GroES to one side of the GroEL double toroid confers a structural asymmetry to GroEL. Understanding the functional implications of this asymmetry requires the ability to distinguish experimentally between the GroES-bound and the GroES-distal rings of GroEL at different stages of the reaction. To achieve this, we took advantage of the observation that

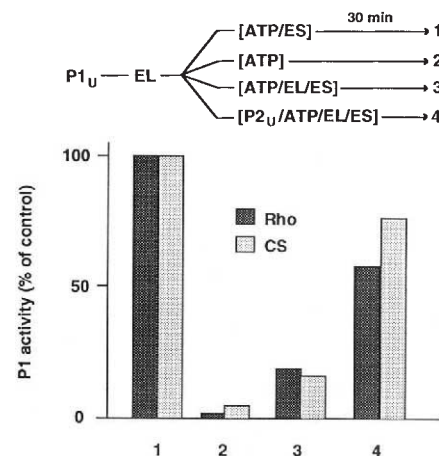
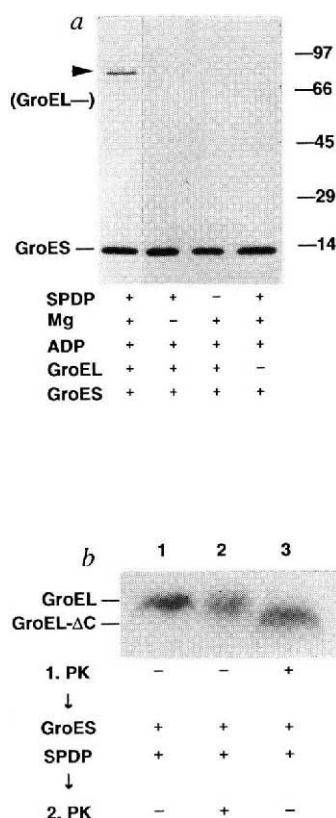


FIG. 2 Exchange of GroES between GroEL–Rho and GroEL–CS complexes during refolding. P1 represents the enzyme whose reactivation is dependent on the transfer of GroES from a complex with GroEL and bound enzyme P2 (see flow diagram, top). The activity of enzyme P1 obtained upon refolding by GroEL/ES is set to 100%. U, unfolded. **METHODS.** Denatured P1 (Rho or CS; prepared as for Fig. 1) was diluted to a final concentration of 1.6 μ M into buffer A containing 0.25 μ M GroEL. After 5 min the sample was split into four aliquots which were diluted 5-fold into buffer A containing 5 mM magnesium acetate, 7.5 mM sodium thiosulphate and 0.5 mM oxaloacetic acid. Reaction 1 received 1 mM ATP and 0.2 μ M GroES; reaction 2, 1 mM ATP; reaction 3, a mixture of 1 mM ATP, 0.2 μ M GroEL and 0.2 μ M GroES, which had been preincubated for 5 min (preformed GroEL–GroES complex); reaction 4, a mixture of 1 mM ATP, 0.2 μ M GroEL and 0.2 μ M GroES which had been preincubated with 1.6 μ M unfolded enzyme P2 (CS or Rho) (preformed GroEL–GroES complex actively refolding enzyme P2). After 30 min incubation, the activity of enzyme P1 was assayed.

FIG. 3 *a*, Covalent crosslinking of GroES and GroEL with SPDP (N-succinimidyl 3-(2-pyridyldithio)propionate). Arrow indicates 70K crosslinked product. *b*, Identification of the GroES-bound toroid of GroEL by protease protection assay.

METHODS. *a*, ^3H -GroES (0.6 μM) was incubated for 30 min at 23 °C with 50 $\mu\text{g ml}^{-1}$ SPDP (added from 4 mg ml^{-1} solution in ethanol) in 20 mM MOPS/KOH, pH 7.2, 50 mM KCl. Remaining amino-reactive groups of the crosslinker were quenched by adding 5 mM glycine in the same buffer. GroEL (0.65 μM) was added with 5 mM magnesium acetate and 1 mM ADP as indicated. After 30 min the reactions were precipitated with trichloroacetic acid. Precipitates were dissolved in SDS-sample buffer (without β -mercaptoethanol) containing 5 mM iodoacetamide to prevent the formation of disulphide bonds and were analysed by non-reducing SDS-PAGE and fluorography. The migration positions of molecular weight markers are indicated *b*, Reaction 1: GroEL-GroES crosslinked product generated as in *a*; reaction 2: after the 30-min incubation with GroEL for crosslinking, an incubation with 20 $\mu\text{g ml}^{-1}$ proteinase K (PK) was performed for 5 min at 4 °C to cleave the 16 C-terminal amino-acid residues from the non-protected subunits of GroEL (2. PK). Unlabelled GroES (6 μM) was present during proteinase K treatment to prevent rebinding and possible crosslinking of free ^3H -GroES that might have dissociated from GroEL. Protease action was stopped by adding 1 mM phenylmethanesulphonyl fluoride; reaction 3: as reaction 1, but GroES was crosslinked to GroEL that had been C-terminally truncated in all 14 subunits (GroEL- ΔC) by a proteinase K treatment before the incubation with GroES (1. PK). Reactions 1–3 were analysed by non-reducing SDS-PAGE and fluorography as in *a*.



GroES protects the subunits of only one of the two GroEL rings from truncation by proteinase K (ref. 8). We had previously estimated that ~50 residues are removed from the C termini of seven of the GroEL subunits. But subsequent mass spectrometry of the resulting GroEL- ΔC subunits has demonstrated that only the 16 C-terminal residues are cleaved (not shown). Notably, this C-terminal truncation does not result in the disruption of GroEL structure nor in a loss in its function^{8,23}.

The observed protease protection indeed occurs in the GroES-bound toroid of GroEL. ^3H -labelled GroES was chemically crosslinked to GroEL before protease treatment using the hetero-bifunctional crosslinker SPDP (Fig. 3 legend) which reacts with amino and sulphhydryl groups. Analysis by SDS-PAGE revealed the presence of a defined, radiolabelled crosslinked product of M_r ~70K, corresponding to the combined mass of one GroEL and one GroES subunit. Formation of this product was dependent on the presence of GroEL and nucleotide (Fig. 3a). As GroES lacks cysteines⁴, one (or more) of the three cysteines of GroEL must participate in crosslinking. When GroES was first coupled to GroEL and then incubated with proteinase K to truncate the subunits in one ring, only full-length GroEL subunits were found to be crosslinked (Fig. 3b). In contrast, coupling of GroES to a GroEL that had been C-terminally truncated in all 14 subunits (GroEL- ΔC) resulted in a faster migrating product, consistent with the size difference

between GroEL and GroEL- ΔC subunits. Thus, when the GroEL-GroES complex is treated with protease, the full-length GroEL subunits define the GroEL ring that binds to GroES.

We used this orientation assay to determine whether GroES binding imposes an asymmetry in the nucleotide occupancy of GroEL. A GroEL-GroES complex was formed in the presence of Mg-ADP and [α - ^{32}P]8N₃-ADP. Azido-ADP and azido-ATP bind to GroEL and upon crosslinking specifically label the nucleotide-binding domain of GroEL²². To enrich for high-affinity-bound nucleotide, GroEL-GroES was subjected to rapid gel filtration under conditions to preserve the complex, and bound 8N₃-ADP was covalently attached by ultraviolet irradiation. Proteinase K treatment was then performed to remove the 16 C-terminal amino acids from the subunits in the GroES-distal ring of GroEL. Labelling with 8N₃-ADP occurred almost exclusively in the protease-protected, GroES-bound GroEL toroid (Fig. 4a). This asymmetric labelling of GroEL was also found when the GroEL-GroES complex was generated in the presence of a mixture of ATP and 8N₃-ATP (not shown). In a separate experiment, analysis of re-isolated GroEL-GroES by scintillation counting and thin-layer chromatography revealed that about seven molecules of ADP were bound to GroEL when the complex was formed in the presence of [^3H]-labelled ATP or ADP (see legend to Fig. 4). In the absence of GroES, almost no bound nucleotide was recovered with GroEL after gel filtration. This confirmed that the labelling with 8N₃-ADP or 8N₃-ATP reflects the nucleotide occupation of the GroEL toroids. The bound but non-crosslinked 8N₃-nucleotides behave essentially like the unmodified forms²². When gel filtration was omitted before crosslinking, covalently coupled 8N₃-ADP was detected in both GroEL rings, although ~80% of the label was in the GroES-associated toroid (Fig. 4a). We conclude that the GroEL-GroES complex has a structural and functional asymmetry. GroES stabilizes the seven subunits of the interacting GroEL ring in a high-affinity state for ADP and, to a lesser extent, favours ADP binding to the opposite ring. Conversely, binding of ADP stabilizes the GroEL-GroES interaction.

Substrate protein-induced ADP release

Does unfolded protein induce the transient release of GroES by altering the nucleotide occupancy of GroEL? GroEL and GroES were incubated with Mg-ATP and [α - ^{32}P]8N₃-ATP as already described, resulting in the formation of a GroEL-GroES complex with seven high-affinity bound nucleotides in the ADP form. Free ATP was removed by gel filtration and aliquots of the reaction mixture received either unfolded rhodanese, citrate synthase or no unfolded protein. Bound 8N₃-ADP was crosslinked after a 30-s incubation. As discussed later, these effects were analysed at different concentrations of K⁺ ions which affect the nucleotide-binding properties of GroEL²⁰. Significantly, the addition of either unfolded protein resulted in a 60–80% reduction of bound ADP, as measured by the decrease in labelling with 8N₃-ADP (Fig. 4b; lanes 1, 2 versus 3–6). Re-isolation of GroEL from these reactions confirmed that the nucleotide had dissociated. Substrate protein also reduced the labelling of the GroEL ring that was not bound to GroES, as seen when the GroEL-GroES complex formed with [α - ^{32}P]8N₃-ADP was not re-isolated by gel filtration before addition of unfolded protein (not shown). These results indicate that binding of substrate protein causes the previously observed stimulation of the GroEL ATPase⁹ by activating ADP-ATP exchange in GroEL.

The bound but not yet crosslinked [α - ^{32}P]8N₃-ADP in the GroES-associated ring of GroEL was not exchangeable with an excess of unlabelled ADP (1 mM) in the absence of substrate protein (Fig. 4b; lanes 7 and 8), even after incubation for up to 5 min. A different effect was noted upon addition of Mg-ATP: within 30 s, 1 mM Mg-ATP exchanged more than 50% of the ADP in the GroES-bound GroEL toroid (lane 10). Under these conditions (with 100 mM K⁺), the GroEL ATPase was ~60% inhibited by GroES to a rate of ~12 ATP molecules per min

(not shown). Complete inhibition by GroES was reached at low K^+ concentrations, however, where the affinity of GroEL for ATP is reduced²⁴ and the ability of GroES to stabilize the ADP state in the opposite GroEL toroid is enhanced. At 1 mM K^+ , ATP addition did not result in the exchange of the tightly bound ADP from GroEL (lane 9), suggesting that ATP hydrolysis may be required for this effect. In contrast, in the presence of substrate protein, efficient dissociation of GroEL-bound ADP was observed at both high and low K^+ concentrations, independent of ongoing ATP hydrolysis in GroEL (lanes 3–6).

Given that the dissociation of ADP induced by substrate pro-

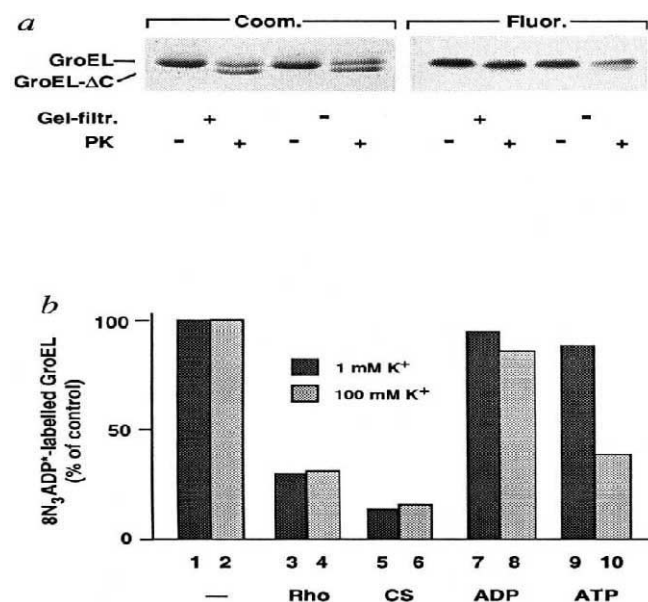


FIG. 4 Interaction of GroEL, GroES and nucleotide analysed by $8N_3$ -nucleotide crosslinking. *a*, High-affinity binding of nucleotide to the GroEL-bound toroid of GroEL. *b*, Dissociation of GroEL-bound $8N_3$ -ADP upon addition of unfolded protein or unlabelled nucleotide. Initially bound and crosslinked [α - ^{32}P]8N₃-ADP is set to 100%. This value corresponds to 7 ADP molecules bound in the GroES-associated toroid of GroEL. The asterisk in the flow-chart indicates α - ^{32}P label.

METHODS. *a*, GroEL (0.15 μ M) and GroES (0.2 μ M) were incubated for 10 min at 23 °C in buffer A containing 5 mM magnesium acetate, 1 μ M [α - ^{32}P]8N₃-ADP (9.47 mCi μ mol⁻¹; ICN)²² and 80 μ M ADP. Half of the reaction was then applied onto a Sephadex-G50 NICK gel filtration column (Pharmacia) equilibrated in buffer A, 5 mM magnesium acetate, 50 μ M ADP to allow for the rapid removal of free and low-affinity-bound [α - ^{32}P]8N₃-nucleotide while preserving the GroEL–GroES complex. After crosslinking by UV irradiation at 312 nm (30 s)²², both reactions were again divided and one half was treated with 20 μ g ml⁻¹ proteinase K for 5 min at 4 °C as for Fig. 3. All samples were analysed by SDS–PAGE, Coomassie-blue staining (lanes 1–4) and fluorography (lanes 5–8). Essentially the same result was obtained when [α - ^{32}P]8N₃-ATP/80 μ M ATP was used in the initial reactions (not shown). When UV crosslinking was omitted, analysis by thin-layer chromatography³¹ showed that the ADP form of the nucleotide was bound to GroEL. *b*, A GroEL–GroES complex was formed in the presence of [α - ^{32}P]8N₃-ATP as in *a*, in 20 mM Tris, pH 7.4, 5 mM magnesium acetate containing either 100 mM KCl or 1 mM KCl, 100 mM NaCl. After gel filtration, both reactions were split into 5 portions each, and either 6 M guanidinium-Cl, unfolded rhodanese (1.2 μ M), unfolded citrate synthase, 1 mM ADP or 1 mM ATP were added as indicated. After a 30-s incubation at 23 °C, the reactions were UV-irradiated to crosslink bound [α - ^{32}P]8N₃-ADP and were analysed by SDS–PAGE followed by fluorography and laser densitometry. Addition of 6 M guanidinium-Cl (columns 1, 2) did not cause any detectable dissociation of bound nucleotide. In a separate experiment, GroEL and GroES were incubated with 0.5 mM [8 - ^{14}C]ATP (52 mCi mmol⁻¹; Amersham) during the initial binding reaction and gel filtration was performed. Analysis by scintillation counting and TLC showed that 6.9 (100 mM K^+ buffer) or 7.2 (1 mM K^+ buffer) ADP molecules were bound per GroEL 14-mer.

tein leads to GroES release, it seemed possible that the ADP–ATP exchange observed in the absence of substrate would have a similar effect. As GroEL and GroES do not separate on a size-exclusion column in the presence of Mg–ATP (Fig. 1), a more sensitive assay was used to test the stability of the GroEL–GroES interaction. As shown previously⁸, a complex of 3H -labelled GroES and GroEL is readily detectable by non-denaturing polyacrylamide gel electrophoresis. Upon incubation with Mg–ADP, 3H -GroES remained bound to GroEL in the presence of up to a 10-fold excess of unlabelled GroES (Fig. 5*a*). But upon incubation with Mg–ATP, a 2-fold molar excess of GroES was sufficient to displace 50% of the bound 3H -GroES, consistent with the possibility that GroES is transiently released from GroEL under these conditions. This assay also confirmed that unfolded rhodanese markedly destabilizes the interaction between GroEL and GroES in the presence of Mg–ADP (Fig. 5*b*).

In conclusion, unfolded protein causes the transient release of GroES from GroEL by reducing the affinity of GroEL for ADP. Likewise, stabilization of the ADP-bound state by GroES destabilizes the interaction of GroEL with substrate protein (J.M. and F.U.H., manuscript in preparation). At cellular concentrations of nucleotides and K^+ , the dissociation of ADP and GroES will result in ATP binding and hydrolysis for substrate protein release.

Alternation of GroES between GroEL rings

Does rebinding of GroES following its release occur at the same or at the opposite end of the GroEL cylinder? GroEL–GroES complex was generated in the presence of Mg–ADP and [α - ^{32}P]8N₃-ADP, and separated from low-affinity bound nucleotide. The complex was irradiated, resulting in the labelling of less than one subunit per GroEL ring that had been bound to GroES at the start of the experiment (Fig. 5*c*; lanes 1, 2). Mg–ATP and unfolded rhodanese were subsequently added for 45 s to allow GroES release and rebinding. Proteinase K treatment was then performed at 4 °C to define the GroEL ring to which GroES was associated at the end of the experiment. All GroEL complexes had GroES bound, as confirmed by the protease protection of half of the GroEL subunits (Fig. 4*a*). Interestingly, we found that the covalently attached [α - ^{32}P]nucleotide was now equally distributed between full-length (GroES-bound) and GroEL–AC subunits of GroEL (Fig. 5*c*, lane 5). This indicates that rebinding of GroES during protein folding can occur to both ends of the GroEL double toroid. The same result was obtained upon incubation of GroEL–GroES with Mg–ATP in the absence of substrate protein, confirming that GroES is indeed transiently released from GroEL under these conditions (lane 4).

Does the binding of substrate protein free GroES to bind to the opposite side of GroEL independently of ATP? To address this question, a GroEL–GroES complex was formed in the presence of Mg–ADP and covalently labelled in the GroES-bound toroid of GroEL with [α - ^{32}P]8N₃-ADP (Fig. 5*c*, lanes 1 and 2). Notably, upon addition of unfolded rhodanese, about 40% of the GroES was found to be reassociated with the opposite end of GroEL, as determined from the appearance of radiolabelled 8N₃-ADP in the truncated GroEL subunits (lane 3). Protease protection assay indicated that all GroEL had GroES bound (results not shown). Taking into account that ~70% of GroEL carried substrate protein, rebinding of GroES to the alternative GroEL toroid seems to be efficient. The substrate-induced rearrangement of GroEL–GroES would normally be accompanied by ATP binding to GroEL.

Discussion

The interaction between GroEL and GroES in the chaperonin reaction cycle is highly dynamic. GroES and unfolded substrate protein counteract each other with respect to binding to GroEL. Whereas GroES stabilizes GroEL in the ADP state, binding of

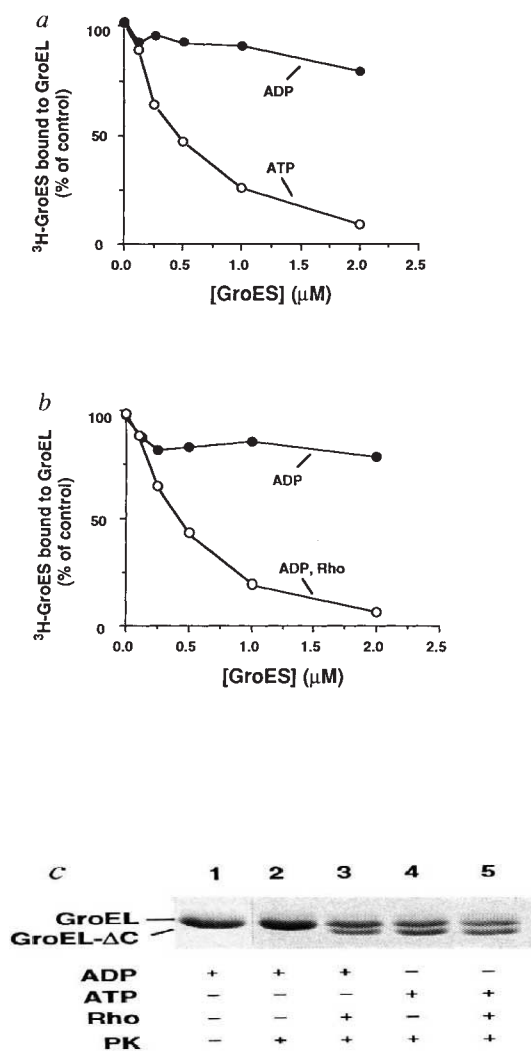


FIG. 5 Exchange of GroEL-bound ^3H -GroES: a, in the presence of ADP (●) or ATP (○), and b, with ADP in the presence (○) or absence (●) of unfolded rhodanese. The amount of initially bound ^3H -GroES is set to 100%. c, Alternation of GroES between GroEL rings during protein folding.

METHODS. a, GroEL (0.25 μM) was incubated at 23 °C in buffer A with ^3H -labelled GroES (0.2 μM), magnesium acetate (5 mM) and either 1 mM ADP or ATP. After 10 min, unlabelled GroES was added at the indicated concentrations, followed by an incubation for 15 min. GroEL-GroES complexes were then separated from the free proteins by non-denaturing PAGE containing the corresponding nucleotide in the polyacrylamide gel and electrophoresis buffers as described⁸. Fluorographs were quantified by laser densitometry. b, GroEL- ^3H -GroES complexes were formed in the presence of 1 mM ADP as in a. Denatured rhodanese (1.4 μM) or the equivalent volume of 6 M guanidinium-Cl was added after unlabelled GroES as indicated. After an incubation for 15 min at 23 °C, analysis by non-denaturing PAGE was performed. c, Formation of GroEL-GroES complexes in the presence of [α - ^{32}P]8N₃-ADP (indicated by 8N₃-ADP* in the flow chart), gel filtration and UV irradiation were performed as for Fig. 4a. After crosslinking to label the GroES-bound toroid of GroEL, the reaction was divided into aliquots, which received 1 mM ADP, 1 mM ATP, unfolded rhodanese (1.2 μM) or both ATP and rhodanese as indicated. After 45 s, all reactions received 10 mM glucose and hexokinase (5 $\mu\text{g ml}^{-1}$; Boehringer) to stop further ATP hydrolysis by GroEL, and incubation with proteinase K (20 $\mu\text{g ml}^{-1}$) was performed for 5 min at 4 °C as for Fig. 3. A fluorograph after SDS-PAGE is shown.

substrate protein causes ADP and GroES dissociation. Importantly, the (re)association of GroES is stimulated by the binding of ATP to GroEL, before ATP hydrolysis triggers the release of substrate protein¹¹. The principle of the reaction is thus defined: both substrate and GroES go through mechanistically coupled cycles of binding and release until the folding protein has lost its affinity for GroEL. A major role of GroES is to coordinate the ATPase of GroEL: regulated by GroES, the subunits in an individual ring will hydrolyse ATP cooperatively^{12,13} either to discharge bound protein or to prepare the toroid for (re)binding of substrate. In addition, the asymmetrical association of GroES with only one ring of GroEL affects the nucleotide- and substrate-binding properties of both GroEL toroids, conferring half-of-the-sites reactivity as recently proposed²⁵.

Our results lead to a model of the GroEL/ES reaction cycle (Fig. 6). In this model: (1) the GroEL-GroES complex is the preferred state of the chaperonin proteins under physiological conditions. ADP is bound with high affinity to the seven subunits of the GroES-associated ring of GroEL. ADP binding in the opposite ring is less tight. Entry of unfolded protein at the GroES distal ring of GroEL triggers ADP dissociation from both GroEL toroids and the release of GroES; (2) in the absence of nucleotide, substrate protein is bound tightly and the affinity for GroES is markedly reduced; (3) upon ATP entry, GroES rebinds at the alternative GroEL ring. ATP binding weakens the interaction between GroEL and substrate protein; (4) simultaneous or rapidly consecutive rounds of ATP hydrolysis in both rings cause release of the bound protein, allowing it to fold in the GroEL cavity; (5) upon substrate dissociation, GroES binding increases in affinity, and rebinding of partially folded substrate in the ADP state continues the folding cycle. Polypeptides that have folded to occlude a sufficient number of their binding sites for GroEL might exit the cavity at this stage. This would also apply to the folded subunits of oligomeric proteins before their assembly. It should be noted that a movement of substrate protein between GroEL rings has not yet been demonstrated. Incompletely folded protein that maintains high affinity for the chaperonin may also rebound to the GroES-attached ring of GroEL. This would trigger a reaction cycle in which GroES does not alternate between GroEL rings.

Binding of GroES to one end of GroEL^{8,18,19} strongly stabilizes the ADP-bound state in the adjacent GroEL subunits and also favours the binding of ADP over ATP in the opposite toroid. This communication between GroEL rings apparently prevents the association of a second GroES⁸. In the absence of substrate and at high concentrations of K^+ and ATP, the GroEL-GroES complex may hydrolyse ATP at about half the rate of free GroEL²⁴ (~12 ATP molecules per min). This ATP hydrolysis is likely to occur in the GroEL toroid opposite GroES and results in ADP-ATP exchange in the GroES-bound ring (not shown in Fig. 6). As a consequence, GroES is transiently released. Assuming that the dissociation of GroES requires the exchange of most of the seven ADP molecules in the adjacent GroEL ring, GroES would cycle slowly at ~1–2 molecules min^{-1} , explaining the apparent stability of its interaction with GroEL in the absence of substrate protein. Binding of unfolded protein triggers ADP release in both GroEL toroids and thus induces the release of GroES independently of ongoing ATP hydrolysis in GroEL. ADP-ATP exchange upon substrate binding may thus occur more or less simultaneously in both GroEL rings (Fig. 6, step 4). GroES will probably then cycle rapidly between bound and free states, consistent with the 5–40-fold stimulation of the GroEL ATPase seen upon substrate binding^{9,13}.

An important question concerns the relative topology of folding protein and GroES during the reaction cycle. Based on electron microscope studies, substrate protein appears to bind within the GroEL cavity at the level of the individual rings^{7,8}. One GroEL ring is thought to provide space for proteins of sizes up to 90K (ref. 7). The unfolded protein may interact initially with

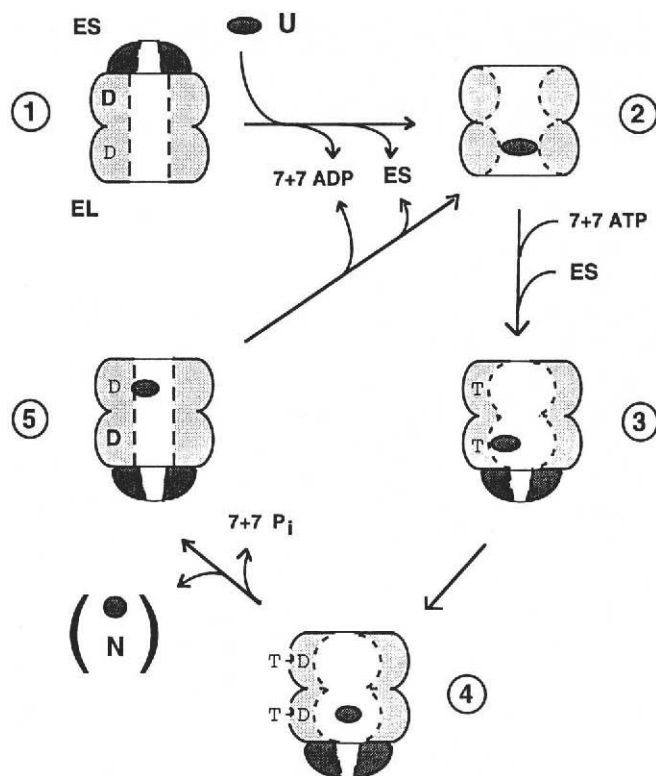


FIG. 6 Model for the reaction cycle of GroEL and GroES in chaperonin-assisted protein folding. D (bold), the high-affinity ADP state in the seven subunits of GroEL which are in direct contact with GroES; D (not bold), the relatively lower ADP affinity of the subunits in the opposite GroEL ring which may hydrolyse ATP dependent on $[K^+]$ and the concentration ratio of ATP/ADP; T, the subunits in a GroEL toroid in the ATP-bound state; U and N, unfolded and native proteins, respectively. Initial binding and rebinding of substrate protein is only shown to the GroEL ring opposite GroES. But the principle of the reaction cycle is considered to be independent of the relative topology of substrate protein and GroES at GroEL. Between steps 5 and 2 GroEL is rotated by 180° , with the substrate protein remaining bound to the same ring. ATP binding and hydrolysis are cooperative events at the level of the 7-subunit rings^{11-13,24}.

the GroES distal toroid of GroEL (Fig. 6, step 1), on the basis of the stability of the GroEL–GroES complex and our observation that induction of the ADP state in GroEL destabilizes the GroEL–substrate interaction. Protein release from a mitochondrial single-ring chaperonin, on the other hand, is dependent on GroES²⁶, and GroES can be crosslinked to GroEL-bound protein²⁷. Apparently, unfolded protein and GroES interact both with the same GroEL ring, probably at the point of substrate release (steps 3 and 4). Consistent with this, the GroEL–GroES complex undergoes a rearrangement upon binding of substrate. GroES is released and can rebind at the opposite ring of GroEL (steps 2 and 3). This ternary complex may represent a ‘transition’ state which relaxes through ATP hydrolysis and substrate release, locking GroEL–GroES into its stable ADP-bound position. ATP binding to GroES itself may contribute to simultaneous occupation of the seven ATP sites in the interacting GroEL ring²².

Unfolded protein is likely to interact with several subunits of a GroEL toroid, presumably through hydrophobic segments exposed in the bound protein. Complete protein release for folding would be due to concerted conformational changes in the GroEL subunits upon cooperative ATP hydrolysis¹⁹ (Fig. 6, steps 3 and 4). Consistent with our model, ATP binding alone is sufficient to resolve the weaker interactions of GroEL with small substrate proteins²⁸, whereas multiple rounds of ATP hydrolysis-dependent release and rebinding seem to be required for the folding of proteins that bind tightly^{9,14,15,21}. The number of cycles of interaction with GroEL necessary for folding⁹ would depend on the folding propensity of a given protein. Assuming that GroES is in contact with the substrate-bound GroEL ring during protein release, the opposite toroid, in the regained ADP state, could trap incompletely folded protein and prevent its premature exit from the GroEL cylinder (steps 4 and 5). □

Received 9 September; accepted 12 October 1993.

- Ellis, R. J. & van der Vies, S. M. *Rev. Biochem.* **60**, 327–347 (1991).
- Gething, M. J. & Sambrook, J. *Nature* **355**, 33–45 (1992).
- Hendrick, J. P. & Hartl, F.-U. *Rev. Biochem.* **62**, 349–384 (1993).
- Hemmingsen, S. M. *et al. Nature* **333**, 330–334 (1988).
- Hendrix, R. W. *J. molec. Biol.* **129**, 375–392 (1979).
- Hohn, T., Hohn, B., Engel, A. & Wurtz, M. *J. molec. Biol.* **129**, 359–373 (1979).
- Braig, K., Simon, M., Furnya, F., Hainfeld, J. F. & Horwich, A. L. *Proc. natn. Acad. Sci. U.S.A.* **90**, 3978–3982 (1993).
- Langer, T., Pfeifer, G., Martin, J., Baumeister, W. & Hartl, F.-U. *EMBO J.* **11**, 4757–4765 (1992).
- Martin, J. *et al. Nature* **352**, 36–42 (1991).
- van der Vies, S. M., Viitanen, P. V., Gatenby, A. A., Lorimer, G. H. & Jaenicke, R. *Biochemistry* **31**, 3635–3644 (1992).
- Bochkareva, E. S., Lissin, N. M., Flynn, G. C., Rothman, J. E. & Girshovich, A. S. *J. biol. Chem.* **267**, 6796–6800 (1992).
- Gray, T. E. & Fersht, A. R. *FEBS Lett.* **292**, 254–258 (1991).
- Jackson, G. S. *et al. Biochemistry* **32**, 2554–2563 (1993).
- Goluboinoff, P., Christeller, J. T., Gatenby, A. A. & Lorimer, G. H. *Nature* **342**, 884–889 (1989).
- Buchner, J. *et al. Biochemistry* **30**, 1586–1591 (1991).
- Landry, S. J., Zeilstra-Ryalls, J., Fayet, O., Georgopoulos, C. & Gierasch, L. M. *Nature* **364**, 255–258 (1993).

- Chandrasekhar, G. N., Tilly, K., Woolford, C., Hendrix, R. & Georgopoulos, C. *J. biol. Chem.* **261**, 12414–12419 (1986).
- Saibil, H. R., Dong, Z., Wood, S. & auf der Mauer, A. *Nature* **353**, 25–26 (1991).
- Saibil, H. R. *et al. Curr. Biol.* **3**, 265–273 (1993).
- Viitanen, P. V. *et al. Biochemistry* **29**, 5665–5671 (1990).
- Mendoza, J. A., Rogers, E., Lorimer, G. H. & Horowitz, P. M. *J. biol. Chem.* **266**, 13044–13049 (1991).
- Martin, J., Geromanos, S., Tempst, P. & Hartl, F.-U. *Nature* **366**, 279–282 (1993).
- McLennan, N. F., Girshovich, A. S., Lissin, N. M., Charters, Y. & Masters, M. *Molec. Microbiol.* **7**, 49–58 (1993).
- Todd, M. J., Viitanen, P. V. & Lorimer, G. H. *Biochemistry* **32**, 8560–8567 (1993).
- Creighton, T. E. *Nature* **352**, 17–18 (1991).
- Viitanen, P. V. *et al. J. biol. Chem.* **267**, 695–698 (1992).
- Bochkareva, E. S. & Girshovich, A. S. *J. biol. Chem.* **267**, 25672–25675 (1992).
- Viitanen, P. V., Donaldson, G. K., Lorimer, G. H., Lubben, T. H. & Gatenby, A. A. *Biochemistry* **30**, 9716–9723 (1991).
- Fayet, O., Louran, J. M. & Georgopoulos, C. *Molec. gen. Genet.* **202**, 435–445 (1986).
- Zhi, W., Landry, S. J., Gierasch, L. M. & Srere, P. A. *Prot. Sci.* **1**, 522–529 (1992).
- Shlomai, J. & Kornberg, A. *J. biol. Chem.* **255**, 6789–6793 (1980).

ACKNOWLEDGEMENTS. We thank J. Ewbank, R. Hlodan and J. Hendrick for discussion, and C. Robinson and S. Radford for mass spectrometry measurements. J.M. is supported by a post-doctoral fellowship from the Dr Mildred Scheel-Stiftung für Krebsforschung, and M.M. is supported by a post-doctoral fellowship from the Human Frontier Science Program.

Geomagnetic field intensity and reversals during the past four million years

Jean-Pierre Valet & Laure Meynadier

Institut de Physique du Globe de Paris, 4 Place Jussieu, 75252 Paris Cedex 05, France

A long and continuous record of the geomagnetic field intensity shows that intensity variations are dominated by two modes. Major episodes of field regeneration prevail on timescales of a few thousand years immediately after most reversals, whereas stable polarity states are characterized by a slow (~ 0.5 Myr) relaxation process. Reversals can be seen as the consequence of a progressive degradation of the dipole field.

SINCE the publication of the first detailed geomagnetic polarity timescale for the past 5 million years (Myr) (ref. 1) no major changes have been introduced in the succession of polarity reversals for this period. The existence of very short polarity intervals is still debated: several such intervals are relatively well documented²⁻⁷ and widely accepted, whereas others are regarded as speculative or artefacts of magnetization. Despite these uncertainties, it is fair to say that changes in field direction are relatively well characterized.

A very different situation exists, however, for records of field intensity. With the exception of recent data sets⁸⁻¹² obtained for the last few hundred thousand years there is no continuous record showing the evolution of field intensity over a long period of time. This is very frustrating, because this parameter is crucial to constrain the mechanisms associated with the regeneration of the geodynamo. Among several aspects, we can emphasize the importance of documenting the long-term variations in the dipole field during stable polarity states, the field evolution across reversals or other geomagnetic events, and the influence of the Earth's orbital parameters such as precession.

The aim of this study is to retrace field intensity variations during the past several million years. To achieve this goal, we need long and continuous sedimentary sequences with a precise control and good resolution in time. Sequences drilled during Leg 138 of the Ocean Drilling Program offered the first opportunity of obtaining a continuous record of the relative changes in field intensity during the past 4 Myr.

The most striking observation from this record is the existence of asymmetrical saw-tooth patterns associated with the succession of the geomagnetic reversals. The geomagnetic intensity decreases slowly during the intervals of stable polarity, but recovers rapidly immediately after a polarity change. The time constants associated with the decay and recovery phases can vary by up to a factor of 100 and are probably governed by the intensity and/or the structure of the convective regime within the Earth's liquid core. Within the limits of the record, the lengths of the polarity intervals seem to depend on the field strength which followed the preceding reversal.

Palaeointensity experiments and results

Sites 848, 851 and 852 of Leg 138 in the equatorial Pacific Ocean were selected on the basis of their detailed magnetostratigraphy^{13,14} and because they offered continuous sequences¹⁵ from at least four holes drilled at each site. A detailed time-depth correlation was obtained by N. Shackleton¹⁶ for successive intervals of ~ 25 kyr after tuning the density variations of the sediment (which reflect the changes in carbonate content) to the insolation curve at 65° N. Measurements were conducted on board R/V *Joides* from the archive halves of the cores. The other measurements were done in the IGP shielded room from U-channels (1 in square section transparent plastic tubes) with the horizontal 2G pass-through cryogenic magnetometer

equipped with high resolution coils. Additional measurements were done on >500 single samples. All these studies involve complete stepwise alternating field demagnetization and analyses of rock magnetic parameters in order to detect changes in the grain sizes and the magnetic mineralogy.

Palaeointensity studies from sediments rely on the assumption that magnetic particles are aligned under the effect of the mag-

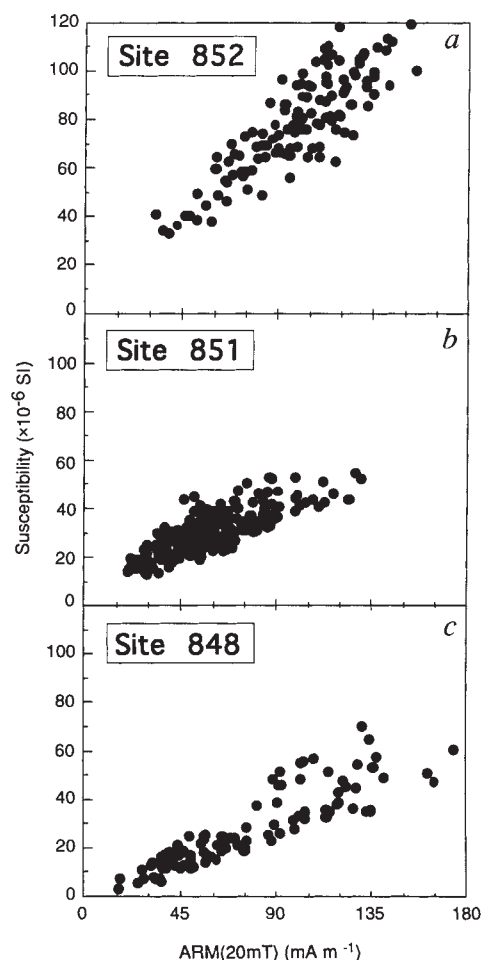
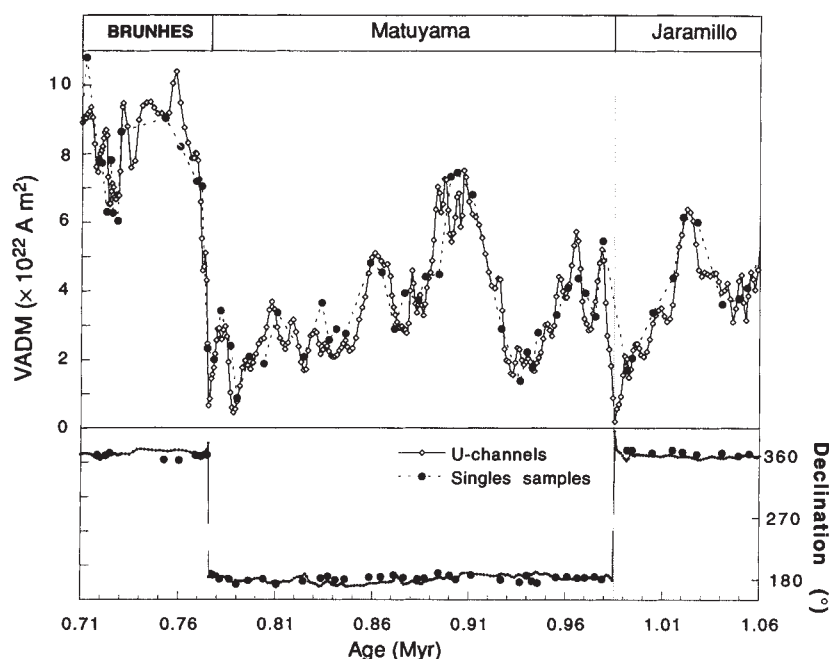


FIG. 1 a-c, Plots of anhysteretic remanent magnetism (ARM) against low-field magnetic susceptibility (K) for single samples measured at three ODP Leg 138 sites from the equatorial Pacific. The linear relationship between these two parameters indicates a uniform magnetic granulometry at every site. ARM (20 mT) indicates that the ARM was determined after demagnetization at 20 mT.

FIG. 2 Declination and relative variations of the dipole field intensity (expressed as VADMs, virtual axial dipole moment, see text) across the Upper Jaramillo and the Matuyama–Brunhes reversals. Note the reproducibility between the measurements of U-channels (measured every 2 cm) and single samples (measured every ~10 cm) obtained from two distinct holes. Very few transitional directions (mostly poorly defined by very weak magnetization intensities) are present.



netic torque exerted by the geomagnetic field so that the magnetic intensity of a sample depends mostly on the field strength and the concentration of magnetic material. Relative field intensities can thus be obtained after normalizing the natural remanent magnetization (NRM) to the amount of magnetic material by using either the anhysteretic remanent magnetization (ARM), the low-field magnetic susceptibility (K) or the isothermal remanent magnetization (IRM). Although the ARM (at the same demagnetization level as the NRM) is usually preferred^{18,19} we wish to stress that magnetic homogeneity within the column of sediment is a fundamental requirement that implies that identical results should be obtained after normalization by any of these three parameters. Another important requirement for reliability of palaeointensity information from sediments is the acquisition of multiple records. These requirements were satisfied for every site. Identical results were obtained with any of the normalization parameters as shown by the linear relationship of the ARM versus K diagrams of the single samples from distinct sites (Fig. 1). There is also a very good reproducibility, not only between distinct holes drilled at every site (Fig. 2), but also between the three distinct sites distributed over 1,000 km. The few intervals with significant changes, mostly due to the loss of magnetic signal, were rejected from the final data set. A detailed summary of these experiments and results is given elsewhere²⁰.

Synthetic palaeointensity record

In Fig. 3, we show our best determination of relative palaeointensity for the past 4 million years (which represents 80 m of sediment) together with the succession of the polarity intervals. As the mean deposition rate does not exceed 2.5 cm kyr^{-1} , we can reasonably assume that the non-dipole components have been averaged out and that the record actually is largely dominated by changes in the intensity of the axial dipole.

Comparisons with other records are extremely limited. Indeed, no other sedimentary record has been published for the period before the Matuyama–Brunhes reversal. The comparison with other records spanning the Brunhes chron⁸ or a part of this chron⁹ is shown elsewhere²⁰ and discussed below.

Relative palaeointensities were converted into virtual axial dipole moments (VADMs) after matching the data with the synthetic record of Meynadier *et al.*¹¹ for the past 140 kyr, which had been calibrated to the volcanic VADMs for that

period^{10,21,22}. The calibration has been tested also with the volcanic data set recently obtained by Schnepf *et al.*^{23,24} for the period 0.45–0.65 Myr before present (BP). Unfortunately, no other detailed volcanic palaeointensities are available for older periods. In the following interpretations, we therefore assume that the downhole response of the sediment to changes in field intensity remained constant. This assumption is supported by the fact that there are no large changes in lithology, mineralogy or deposition rates.

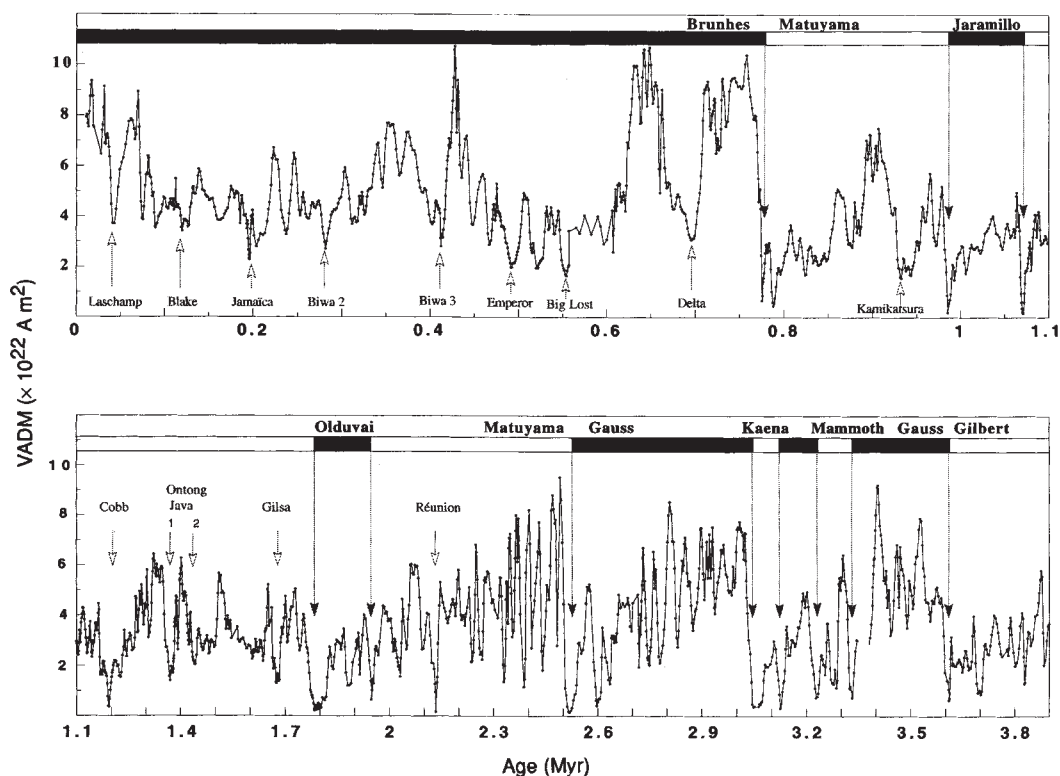
The mean VADM of $3.9 \pm 1.9 \times 10^{22} \text{ A m}^2$ is not significantly different from the values calculated for the past 140 kyr ($5 \pm 2 \times 10^{22} \text{ A m}^2$) or the period between 15 and 50 kyr BP ($4.3 \pm 1.5 \times 10^{22} \text{ A m}^2$) from sedimentary and volcanic data sets respectively^{11,12}. It differs, however, by almost a factor of two from the mean VADM deduced from volcanic data for the past 5 Myr (ref. 25). This discrepancy could be partly explained by the relatively low number of absolute palaeointensities (only 150

TABLE 1 Ages of the recent geomagnetic events

Event	Age (Myr BP)	Age	Difference (Myr)
	Champion <i>et al.</i> (ref. 4)	(Myr BP) This work	
Laschamp	0.042 ± 0.010	0.040	−0.002
Blake	0.114 ± 0.001	0.118	+0.004
Jamaica	0.182 ± 0.031	0.195	+0.013
Biwa II	0.289 ± 0.019	0.280	−0.009
Biwa III	0.389 ± 0.009	0.412	+0.022
Empereur	0.443 ± 0.019	0.419	+0.049
Big Lost	0.565 ± 0.014	0.554	−0.011
Delta	0.635 ± 0.005	0.690	+0.055
Matuyama–Brunhes	0.730	0.780	+0.050
Kamikatsura	0.850 ± 0.030	0.931	+0.081
Jaramillo (upper)	0.910	0.990	+0.080
Jaramillo (lower)	0.980	1.070	+0.090
Cobb Mountain	1.100	1.190	+0.090

Comparison between ages of the excursions, short events, and most recent reversals of the geomagnetic field deduced from the compilation of Champion *et al.*⁴ and ages of the intensity lows observed in this work. The offset between the two ranges of dates for ages >0.62 Myr is caused by the recent recalibration of the polarity timescale^{20,41}. The differences between the two sets of ages are not larger for excursions and short events than for the reversals.

FIG. 3 Relative variations of the dipole field intensity during the past 4 Myr obtained from a composite sequence from ODP Sites 848 and 851. The polarity intervals are indicated by horizontal bars (black/white shows normal/reverse polarities) and the position of the reversals shown by solid arrows. The excursions and the short events observed in previous studies are indicated by open arrows^{4,26-28} and correlated with intensity minima of the present record. In order to have a detailed view of the Brunhes chron, the upper and lower figures are not plotted on the same horizontal scale.



determinations) without precise ages in many cases and therefore no control on the regularity of their temporal distribution within this period. The present record indicates also that there is no significant difference between the mean dipole field intensities of the normal and reverse polarities ($4.1 \pm 2.0 \times 10^{22} \text{ A m}^2$ and $3.3 \pm 1.5 \times 10^{22} \text{ A m}^2$, respectively), given the rather large standard deviations and the fact that the mean normal VADM is slightly biased by the presence of high intensities during the Brunhes chron. We thus prefer to interpret this result as an indication that the two polarity states are equally stable.

The period from 2 Myr BP to the present is characterized by several intensity drops of variable duration which are not correlated with field reversals (Fig. 3). In many cases their occurrences coincide with the ages of geomagnetic field excursions (defined as sudden directional changes followed by a return to the initial polarity) and/or short events (that is, extremely short polarity intervals) shown with open arrows in Fig. 3 which have been reported in several studies for the period 0–1.1 Myr BP and summarized by Champion *et al.*⁴ (Table 1). They agree also with recent observations and datings of the Gilsa²⁶, Cobb Mountain²⁷ and Ontong Java 1 and 2 events²⁸ for times before 1.1 Myr BP. Only one large deviation from dipolar directions is observed in the present record of the Cobb Mountain event. The interpretation of transitional directions in terms of short geomagnetic features in weakly magnetized sediments has often been questioned. The observation of field intensity decreases and the fact that intermediate or reverse directions^{4,29} have been reported from volcanic sequences for at least some of these events suggest that they could be real geomagnetic features. We notice, however, that other intensity drops are present and not associated with any known event. Additional detailed records would be necessary to clarify further this point.

Saw-tooth intensity changes

The intensity pattern associated with field reversals can be best described as an asymmetric saw-tooth curve (Figs 2 and 3) with a gradual decrease before the transitions and a very rapid recovery immediately following the directional changes. The overall pat-

tern of the record could recall typical numerical solutions of two-disc dynamo systems³⁰⁻³². None of these models, however, predict large quasi-linear intensity decreases but rather an increase (or a decrease) in the amplitude of the oscillations before a reversal. The fluctuations observed in the present record do not show evidence for such amplitude changes but their distribution could suggest the existence of short periodicities. Spectral analyses have therefore been performed on the signals of the rock magnetic parameters (NRM(20 mT), defined as the NRM after demagnetization at 20 mT, ARM, K and ARM(20 mT)/ K) and those of relative palaeointensity NRM(20 mT)/ARM(20 mT), NRM(20 mT)/ K . The main common characteristic feature is the absence of a stationary

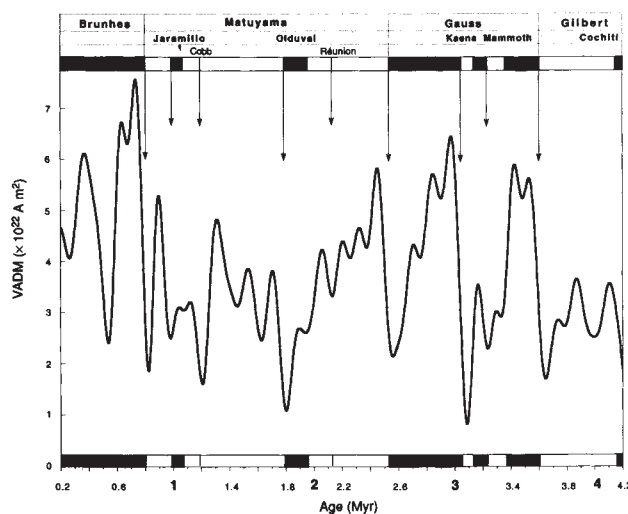


FIG. 4 Record of the dipole field intensity obtained after filtering out the components with periods < 125 kyr. The curve depicts the successive episodes of intensity loss and recovery and their relationship to reversals.

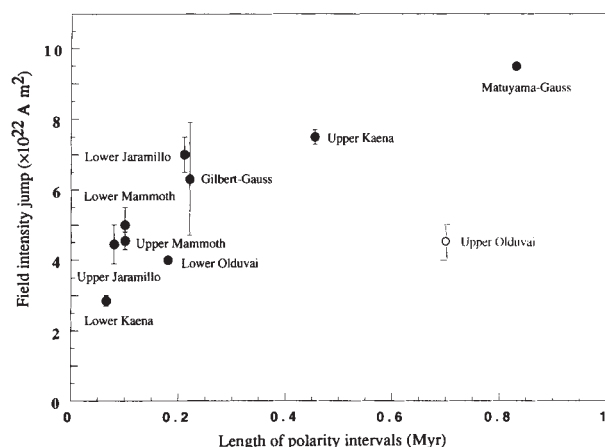


FIG. 5 Relationship between the amplitude of the dipole field following the reversals and the length of the subsequent polarity intervals. The field strength following the upper Olduvai does not fit with the overall tendency (see text).

signal for periods larger than 50 kyr with different frequency spectra for different periods. Although there is no correlation between the rock magnetic parameters and the ratios indicative of relative field intensity, the possible influence of climatically controlled factors should not be underestimated. So far, no appropriate technique is available to detect and extract short-term variations induced by these phenomena. Studies are currently in progress on this aspect.

We can, however, remove most of the energy in the spectra potentially related to climatic modulation by filtering out the higher frequencies (that is, periods <125 kyr of the signal). After this process the 400-kyr Milankovitch component is thus the only climatic periodicity which was not removed. The results shown in Fig. 4 have the advantage of displaying a global aspect of the record and do not show any dominance or superposition of a 400-kyr fluctuation. The first striking feature is that some reversals correspond to particularly large and rapid recovery. Such is the case for the three major reversals that separate the original geomagnetic 'periods': Gilbert-Gauss, Gauss-Matuyama and Matuyama-Brunhes. The intensity recovery occurring at the end of the Kaena subchron appears to be of equal importance. Other reversals are marked by smaller amplitude fluctuations such as the lower and upper Mammoth and Jaramillo. Some other reversals are not associated with a very striking intensity behaviour in Fig. 4: lower Kaena, lower Olduvai or lower Jaramillo; however, those boundaries which are beginnings of short subchrons with typical durations of the order of 100 kyr are in fact accompanied by a smaller intensity recovery which has been smoothed out by the process of filtering and can be clearly observed in Fig. 3. Finally, some short-term events are associated with significant fluctuations which are also more visible in Fig. 3: Cobb Mountain, Ontong Java 1 and 2, Gilsa and Réunion. Three rather regular 'cycles' follow major recoveries at Gilbert-Gauss, upper Kaena and Gauss-Matuyama. The cycle is, however, not so clear for the period following the upper Olduvai with rather a pseudo-reverse saw-tooth pattern, but it happens to be more obvious in new, still unpublished, results of another core from the Indian Ocean (MD940). This could be explained by changes in the magnetic homogeneity of the sediment reflected by larger changes in the rock magnetic properties. Indeed, it is striking that the palaeointensity signal was limited to Site 848 for this period and was not duplicated at other sites²⁰.

Although it does not seem to have been stressed before, the saw-tooth pattern can be taken as a general characteristic of the geomagnetic field. Indeed, if we include the new results of the upper Olduvai, this observation holds for 90% of reversals

covered by the present data set. In addition, the same intensity pattern across the Matuyama-Brunhes reversal has now been observed in four distinct marine cores distributed around the world³³. Similar asymmetry between the decay and recovery phases can be observed in the few detailed palaeointensity records of reversals. Two typical examples are given by the sedimentary records from Crete³⁴ (6 Myr BP) and the reversal from the Steens Mountain³⁵ (15 Myr BP) which is the most detailed palaeointensity record obtained so far from a volcanic sequence. A recent study of lava flows from Kauai³⁶ concludes that the post-transitional lava flows are characterized by stronger palaeofield intensities than the pre-reversal period. Therefore, this observation seems to be an overall and characteristic feature of intensity variations accompanying field reversals.

Despite the presence of a major field recovery after the last reversal, the Brunhes chron seems to be characterized by a more complex behaviour than the rest of the record. Indeed, the very few studies covering this interval published so far^{8,9} do not show evidence for a regular decrease of field intensity since the last reversal. Ironically, we noticed that no major polarity reversal has been observed yet since the last one (0.78 Myr BP). It is more appropriate to wonder whether the palaeointensity records are still valid within the first few metres of sediment that are characterized by high water content and thus more subject to physical disturbances and subsequent reorientations of the magnetic grains. Kodama and Sun³⁷ recently suggested that compaction of sediment can eventually induce an intensity decrease during inclination shallowing. It is possible that the segment of core covering the Brunhes period was not significantly compacted. Although such mechanisms should require a larger amount of clay than in the present cores, the average values of the inclinations (I) of the successive subchrons (for example, $I=2.2^\circ$ for Brunhes, $I=-4.9^\circ$ from the upper Jaramillo to Matuyama-Brunhes, $I=3^\circ$ for Jaramillo, $I=-6^\circ$ before Jaramillo) do not show evidence for shallowing in the older parts of the cores.

Characteristic times

The typical time constants of the saw-tooth cycles range between 0.5 Myr and 1 Myr in Fig. 4 (recall that periods <125 kyr have been filtered out) and seem to correspond to some kind of relaxation process. These time constants are significantly longer than the slower modes of free diffusion of the field and could therefore be related to the free decay of the convective core regime itself.

Recovery times are smoothed in Fig. 4 and so must be estimated from the original record of Fig. 3. The best determinations for intervals with high deposition rates indicate a field change between 1 and $2 \times 10^{22} \text{ A m}^2 \text{ kyr}^{-1}$, that is similar to the decrease of the dipole field over the past 2,000 yr. If effects due to smoothing of the signal induced by post-depositional reorientation of magnetic grains are taken into account, these rates could be as high as $8 \times 10^{22} \text{ A m}^2 \text{ kyr}^{-1}$ (assuming a blocking depth of magnetization between 10 and 15 cm (refs 38, 39)). Such values correspond to time constants of 1,000 years only, that is much shorter than diffusion times and therefore necessarily reflect the dynamics of regeneration in the dynamo process. Therefore, we observe a scale factor of about 100 between the extreme time constants associated with the two main phases inducing the high asymmetry of the saw-tooth features.

The pattern of Fig. 4 supports the major chron boundaries. It seems that some reversals are so 'powerful' that field intensity is regenerated for a long time and that the decaying phase can eventually include short events or excursions which have only second or third order consequences for amplitude fluctuations. This would imply that the core kept some memory of these powerful reversals. Within the limits in resolution of the record, we can establish a rough correlation between the amplitude of intensity recovery following the reversals and the duration of the subsequent polarity chrons or subchrons (Fig. 5); the only significant deviation from this relationship is the upper Olduvai but, as we mentioned it before, does not seem to be observed in

other data. Therefore, the stability of the polarity intervals seems to depend on the amplitude of field intensity jump which followed the previous polarity transition. If, for some reason a reversal is not followed by immediate large field recovery, there is no stable antipodal polarity and the field reverses again immediately (observed typically after the first swing or directional change accompanying short events and excursions). Conversely a large field recovery will generate a long stable polarity interval. Similar conclusions have been reached on theoretical grounds by Olson and Hagee⁴⁰ who suggest that small changes in the intensity or the structure of the outer core convection can result

in an irregular pattern of reversals. The nonlinear kinematic dynamo calculations of these authors predict a correlation between field strength and epoch length.

The selective and tantalizing pattern that emerges from our measurements requires additional work and subsequent confirmation (partly obtained already) from marine and volcanic sequences in other parts of the world. Analyses of the Brunhes period remains critical; it is also important to analyse records spanning periods older than the past 4 Myr. If confirmed, our results are likely to provide major constraints to the competing dynamo theories. □

Received 4 June; accepted 12 October 1993.

1. Cox, A. & Dalrymple, G. B. *Earth planet. Sci. Lett.* **3**, 173–177 (1967).
2. Verosub, K. *Rev. Geophys. Space Phys.* **15**, 129–143 (1977).
3. Hoffman, K. A. *Nature* **294**, 67–68 (1981).
4. Champion, D. E., Lanphere, M. A. & Kuntz, M. A. *J. geophys. Res.* **93**, 11667–11680 (1988).
5. Lund, S. et al. *Geophys. Res. Lett.* **15**, 1101–1104 (1988).
6. Clement, B. M. & Constable, C. G. *Rev. Geophys. suppl. (US nat. Rep. IUGG)* 433–442 (1991).
7. Tric, E. et al. *Phys. Earth planet. Inter.* **63**, 319–336 (1991).
8. Kent, D. V. & Opdyke, N. D. *Nature* **266**, 156–159 (1977).
9. Tauxe, L. & Wu, G. J. *geophys. Res.* **95**, 12337–12350 (1990).
10. Tric, E. et al. *J. geophys. Res.* **97**, 9337–9351 (1992).
11. Meynadier, L., Valet, J.-P., Weeks, R., Shackleton, N. J. & Hagee, V. L. *Earth planet. Sci. Lett.* **114**, 39–57 (1992).
12. Thouveny, N. et al. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **7**, 157–172 (1993).
13. Mayer, L. et al. *Proc. ODP Init. Repts No. 138* (Ocean Drilling Program, College Station, Texas, 1993).
14. Schneider, D. A. *Proc. ODP Sci. Res.* **138** (in the press).
15. Hagelberg, T. et al. *Proc. ODP Init. Repts No. 138*, 79–85 (Ocean Drilling Program, College Station, Texas, 1992).
16. Shackleton, N. J. et al. *Proc. ODP Sci. Res.* (in the press).
17. Nagy, E. A. & Valet, J.-P. *Geophys. Res. Lett.* **20**, 671–674 (1993).
18. Levi, S. & Banerjee, S. K. *Earth planet. Sci. Lett.* **29**, 219–226 (1974).
19. King, J. W., Banerjee, S. K. & Marvin, J. J. *geophys. Res.* **88**, 5911–5921 (1983).
20. Meynadier, L., Valet, J.-P. & Shackleton, N. J. *Proc. ODP Sci. Res.* **138** (in the press).
21. McElhinny, M. W. & Senanayake, W. E. *J. Geomagn. Geoelect.* **34**, 39–51 (1982).
22. Coe, R. S., Gromme, S. & Mankinen, E. A. *J. geophys. Res.* **83**, 1740–1756 (1978).
23. Schnepf, E. *Phys. Earth planet. Inter.* **70**, 231–236 (1992).

24. Schnepf, E. *Geophys. J. Int.* (in the press).
25. McFadden, P. L. & McElhinny, M. W. *J. Geomagn. Geoelect.* **34**, 163–189 (1982).
26. Clement, B. M. & Kent, D. V. *Earth planet. Sci. Lett.* **104**, 48–58 (1987).
27. Clement, B. M. & Martinson, D. G. *J. geophys. Res.* **97**, 1735–1751 (1992).
28. Gallet, Y., Gee, J., Tauxe, L. & Tarduno, J. G. *Proc ODP Sci. Res.* **130**, 547–559 (1993).
29. Roperch, P. & Duncan, R. A. *J. geophys. Res.* **95**, 2713–2726 (1992).
30. Rikitake, T. *Proc. Camb. Phil. Soc. math. phys. Sci.* **54**, 89 (1958).
31. Cox, A. *J. geophys. Res.* **73**, 3247–3260 (1968).
32. Nozières, P. *Phys. Earth planet. Inter.* **17**, 55–74 (1978).
33. Valet, J.-P., Meynadier, L., Bassinot, F. & Garnier, F. *Geophys. Res. Lett.* (in the press).
34. Valet, J.-P., Laj, C., Langereis, C. G. *J. geophys. Res.* **93**, 1131–1151 (1988).
35. Prévot, M., Mankinen, E. A., Gromme, C. S. & Coe, R. S. *J. geophys. Res.* **90**, 10417–10448 (1985).
36. Bogue, S. W. (*abstr.*) *EOS* **74**(16), 109 (1993).
37. Kodama, K. P. & Sun, W. W. *Geophys. J. Int.* **111**, 465–469 (1993).
38. Tucker, P. *Phys. Earth planet. Inter.* **20**, 11–14 (1979).
39. Tucker, P. *Geophys. J. R. astr. Soc.* **63**, 149–163 (1980).
40. Olson, P. & Hagee, V. L. *J. geophys. Res.* **95**, 4609–4620 (1990).
41. Shackleton, N. J., Berger, A. & Peltier, W. R. *Trans. R. Soc. Edinb., Earth Sci.* **81**, 251–261 (1990).

ACKNOWLEDGEMENTS. We thank V. Courtillot for reviewing the manuscript, and for discussions, D. Jault for help, J. Shackleton for providing a detailed and accurate timescale and E. Nagy for developing software for the cryogenic magnetometers. Samples and U-channels were obtained by permission of the Ocean Drilling Program. The manuscript benefited from reviews by L. Tauxe and J. Gien. Support for this study was provided by the French CNRS-INSU-ATP géosciences marines.

Geomagnetic palaeointensities during the Cretaceous normal superchron measured using submarine basaltic glass

Thomas Pick* & Lisa Tauxe*

Scripps Institution of Oceanography, La Jolla, California 92093, USA

High-quality palaeointensity data have been obtained from Thellier–Thellier experiments on recent and Cretaceous submarine basaltic glasses. Whereas the recent samples faithfully yield today's geomagnetic intensity at the site, the palaeointensities for the beginning and end of the Cretaceous normal superchron are only 45% and 25%, respectively, of today's value. The data thus extend the 'Mesozoic dipole low' into the Cretaceous superchron, and confirm that submarine basaltic glass is an excellent material for palaeointensity studies.

LONG-TERM changes in the frequency of polarity reversals of the geomagnetic field are intriguing, but poorly understood, signals of the deep interior of the Earth. Twenty-five years ago, Cox speculated¹ that these changes might reflect long-term changes in the intensity ratio of dipole and non-dipole fields. He suggested that such intensity variations would probably be caused by changes in core–mantle boundary conditions. Most modern

studies agree that variations in the thermal boundary conditions imposed on the core by the mantle could alter the fluid flow in the liquid outer core and ultimately change the reversal frequency^{2,3}. The mechanism that ties processes at the core–mantle boundary (CMB) to changes in the geomagnetic field, and the exact consequences of these variations, are not well understood and remain the subject of considerable disagreement and speculation. Indeed, the two predominant models that link CMB processes to reversal frequency come to diametrically opposite conclusions. Larson and Olson⁴ propose that the fre-

* Present address: Fort Hoofddijk Paleomagnetic Laboratory, University of Utrecht, The Netherlands.

quency of geomagnetic reversals is inversely related to the vigour of core convection. They argue that a thin CMB would increase the flow of heat from the outer core to the mantle and so create more vigorous convection in the outer core, resulting in a stable and strong non-reversing field. A thick boundary layer, on the other hand, would reduce the core convection and increase the reversal frequency. In contrast, Loper and McCartney⁵ predict that increased convective vigour in the core would increase the reversal rate by generating frequent instabilities. A key piece in this puzzle of the coupling of core, mantle and surface events is the Cretaceous normal superchron (CNS), marking the time period from 124 to 84 Myr BP, during which there was constant normal polarity of the geomagnetic field⁶. In his 'superplume' scheme, Larson⁴ reasons that plume formation directly above the CMB (also known as the D'' layer) would increase convection in the outer core and that the large amount of heat lost would produce a strong, stable, non-reversing dipole field, that is, a geomagnetic polarity superchron. Courtillot and Besse⁷ (among others^{5,8}) argue that superchrons represent a low-energy state with infrequent instabilities, and thus correspond to inactive periods in the D'' layer, when it is thick and heat transfer across the CMB is relatively small. Both models are intuitively reasonable; that of Larson and Olson⁴ is testable in that it predicts an inverse correlation between field intensity and reversal rate. Knowledge of the palaeointensity of the geomagnetic field during the CNS would therefore provide key evidence in this debate.

We present here new results from Thellier-Thellier experiments on recent and Cretaceous submarine basaltic glasses (SBG) from the 'BBQ'-Site and from four Deep Sea Drilling Project/Ocean Drilling Programme (DSDP/ODP) holes, respectively. The palaeointensities derived from the recent glass reproduce precisely the present field at the site. This, together with the high-quality results obtained from the Cretaceous samples, demonstrates that SBG is an excellent material for palaeointensity studies. SBG is widely distributed throughout the oceans (in a range of ages) and therefore holds great potential for future palaeointensity studies.

The results from the Cretaceous glasses extend the 'Mesozoic dipole low'⁹ into the CNS. These low intensities at the beginning and end of the superchron do not support the existence of a strong magnetic field during the Cretaceous as proposed by Larson's⁴ 'superplume' model. Instead, they support the notion of Prévot *et al.*⁹ that the processes controlling reversal frequency and palaeointensity are probably decoupled.

Palaeointensity problematics

Palaeointensity data are uncommon, both in age and spatial distribution, and become extremely scarce before the Holocene. This is because most of the igneous rocks used for palaeointensity studies alter during the experiment, oxidizing the magnetic minerals.

In their updated palaeointensity record for the Mesozoic and Cenozoic eras, Prévot *et al.*⁹ show evidence for long-term variations, with intensities being significantly lower between 175 and 125 Myr than during the Cenozoic. Only two data sets from the CNS are available^{10,11}, and their reliability can be questioned. The first stems from the Coniacian-Santonian boundary at the end of the superchron¹⁰. Although the data have been obtained from baked sediments, they do not provide a conclusive baked contact test. Thus, there is no strong evidence that the magnetization is due to baking. Moreover the age of the magnetization is poorly constrained and may not be CNS at all. In a recently published study on Cretaceous volcanic rocks from Israel and India, Sherwood *et al.*¹¹ show that their samples were subject to severe alteration during the palaeointensity experiment. Despite the fact that practically none of the data meet Prévot *et al.*'s⁹ selection criteria, a great number were used for palaeointensity calculations. Further points of concern include the subjective way in which intensity data were selected as well as poorly constrained ages.

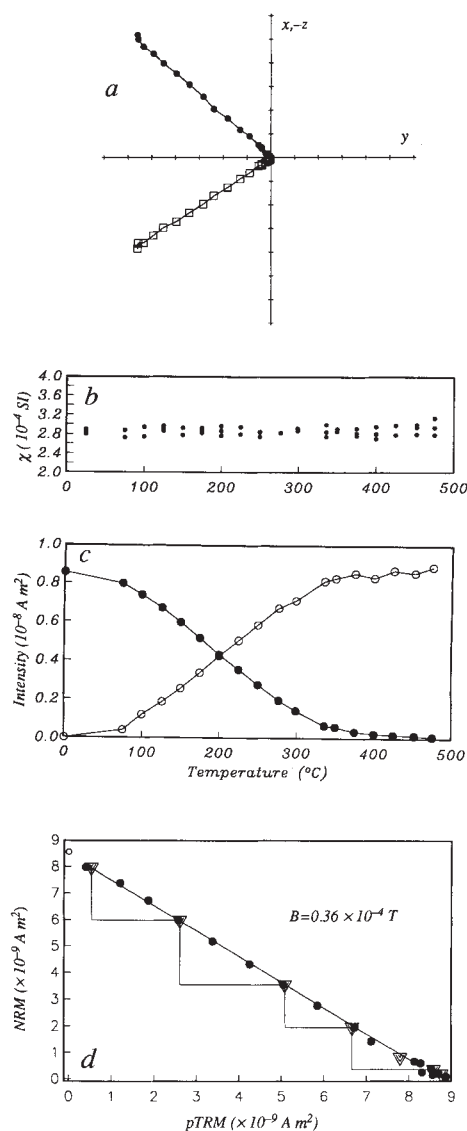


FIG. 1 Thellier-Thellier experiment for BBQ sample 23522e. a, Zijderveld diagram of the thermal demagnetization data of the sample. Filled circles and open squares indicate the horizontal and vertical components of the magnetization, respectively. The sample shown here has a well-defined characteristic component, as do all the other samples. A small, low-temperature component is evident and was explained as a viscous magnetization, acquired during sample preparation¹³ (tick-marks on the axes represent $1 \times 10^{-9} \text{ A m}^2$). b, Magnetic susceptibility (χ) plotted against temperature. Changes in χ can indicate alteration of the magnetic minerals. About 90% of the BBQ samples show no changes up to temperatures above 500 °C. c, Thermal demagnetization of the natural remanence (NRM) (the original TRM) shown as filled circles and the acquisition of the laboratory-induced partial thermal remanence (pTRM, open circles) as a function of temperature. The intersection of both curves lies at exactly 50% of the initial NRM, indicating a single domain remanence carrier³⁴. d, Alternative way of plotting the data from c as a so-called Arai plot¹⁴: NRM-decay versus the pTRM acquisition (filled circles). If a sample does not alter during the experiment, the NRM-pTRM ratio should remain constant throughout the experiment, and the data should lie along a straight line of negative slope. Provided that the TRM is linearly related to the field, the absolute value of the slope of the line equals the ratio of ancient field to laboratory field. The so-called 'pTRM-checks'¹² (triangles) are derived from a second pTRM, acquired at a given temperature, after the sample has been demagnetized at a significantly higher temperature. If the sample did not alter, it will acquire the same pTRM a second time and this plotted against the original NRM should fall on the line of initial data (filled circles). Deviations, on the other hand, could suggest sample alteration. (B indicates the palaeointensity calculated from this sample.)

The BBQ palaeointensity experiment

We have recently presented data from Thellier–Thellier¹² experiments on Holocene SBG from the East Pacific Rise (EPR) that have shown that this is potentially the ‘ideal’ material for absolute palaeointensity determinations¹³. These Holocene samples also contained ‘zero age’ glasses which reproduced the present field at the sample location (known from the International Geomagnetic Reference Field, IGRF). The age of these glasses had, however, been radiometrically determined with an error of ± 500 yr. Naturally, as the intensity of the geomagnetic field has changed by roughly a factor of 2 in that time, the agreement of those samples with the IGRF could be fortuitous. Therefore we find it appropriate to emphasize the suitability of this material for Thellier–Thellier¹² experiments by discussing some additional results from recent glasses obtained from the widely reported BBQ-Site at 9°50' N on the EPR.

Submarine volcanics of mid-ocean-ridge basalt (MORB) composition were extruded on the ocean floor at the so-called BBQ-Site while scientists watched through the porthole of the deep-sea submersible *Alvin* in 1990. The researchers marked the location and went back one year later to sample the sheet flow. This ‘zero age’ material is uniquely suited for testing the validity of using SBG for palaeointensity studies, as both the age of the glass (from direct observation) and the intensity of the geomagnetic field at the site at the time of generation, and thus the expected answer (from the IGRF), are exactly known.

Figure 1 shows the results of a Thellier experiment for a representative sample. In Fig. 1a we plot a vector end-point diagram for thermal demagnetization, indicating a well defined characteristic component with a small, low-temperature overprint due to sample preparation¹³. We also plot susceptibility (χ) in Fig. 1b as these can be an indication of sample alteration. None was observed. Figure 1c shows the decay of natural remanence (NRM, which is the original thermal remanence, shown as filled circles) and acquisition of the laboratory-induced partial thermal remanence (pTRM, shown as open circles). In the so-called Arai plot¹⁴ (Fig. 1d), we show the NRM-decay against the pTRM acquisition (filled circles). A straight line (as shown here) indicates that samples have not altered during the experiment; another indication of this lack of alteration is that ‘pTRM-checks’¹² lie on or close to the line. The palaeointensity of the sample can be calculated from the slope of the line (see Fig. 1 legend for further information on pTRM-checks and palaeointensity calculations).

All 12 BBQ samples performed exceptionally well. The palaeointensity calculations show a mean intensity of 35 μ T with a standard deviation of 2 μ T. The expected field for this location as calculated from the IGRF is 37 μ T. The palaeointensity derived from the BBQ glass result is therefore well within the expected experimental error and supports our earlier claim that SBG is a suitable material for Thellier experiments.

Glasses from the Cretaceous superchron

Because basaltic glass is often recovered in both dredged and drilled material from the ocean floor, the availability of submarine basaltic glass holds great potential for achieving a better distribution of palaeointensity data through time and over the Earth's surface. The real challenge of this study was to determine whether SBG that was older than the Holocene would also be suitable for palaeointensity experiments. We selected glass from near the beginning (Holes 417D, 418A and 807C) and the end (Hole 543A) of the CNS. Holes 417D and 418A are located in the Caribbean, and Hole 807C lies on the Ontong Java Plateau in the southeast Pacific Ocean. They are all of Early Aptian age and span the onset of the CNS. Hole 543A is located east of the Lesser Antilles in the Atlantic Ocean and is of Campanian age, just before the end of the CNS.

Whereas Holes 417D and 418A are of M0 (where M0 is the youngest member of the M-sequence of sea-floor magnetic anomalies) age, Holes 807C and 543A lie just inside the CNS.

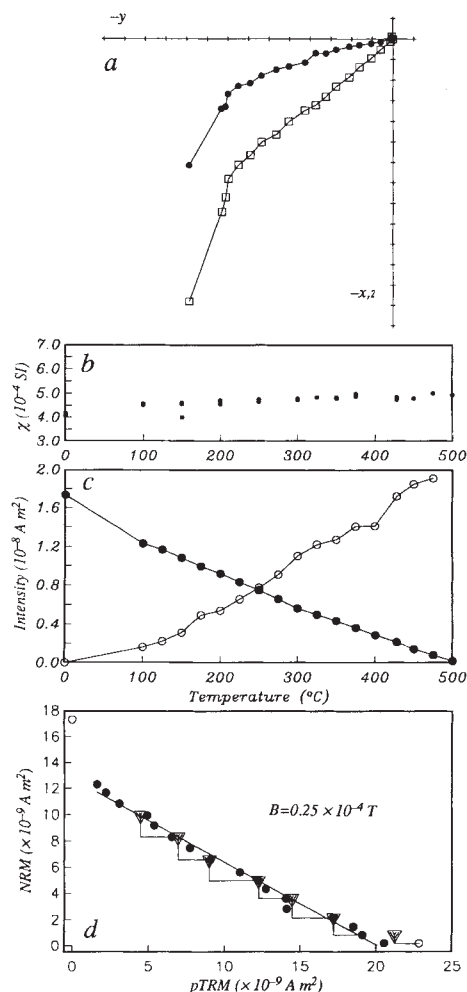


FIG. 2 Same as Fig. 1 for representative ‘CNS’-sample 807c931p7a. Note the larger contribution of a low-temperature viscous component, excluding data below 200 °C from palaeointensity calculation.

Although $^{40}\text{Ar}/^{39}\text{Ar}$ ages are available for Holes 418A and 807C, the absolute age of M0 is still a subject of discussion. Whereas the whole-rock age for the basement of 807C (earliest CNS) gives a well behaved plateau age of 122.3 ± 1.0 Myr¹⁵, the date of 125.8 Myr for 418A (M0) is derived from a highly distorted age spectrum¹⁶. Most timescales^{17,18} use an age of 118 Myr for the end of M0 based on Harland *et al.*¹⁹. Because the relative age of the older, reverse-magnetized rocks from 417D and 418A and the younger 807C samples (of normal magnetization) is well established, an age of 118 Myr (which had considerable uncertainties¹⁹), is clearly too young. The new 1989 chronogram age of 125.5 Myr for the Barramian/Aptian boundary⁶ is also of little help because of its large uncertainty. However, because the age of 124 Myr for the base of the CNS (based on two tie points at the Aptian/Albian boundary (112 Myr) and the Hauterivian/Valanginian boundary (135.5 Myr)) and the 807C-age of 122.3 Myr seem well established, we have used a nominal age of 125 Myr for Holes 418A and 417D and hence for M0. Hole 543A was drilled on the extreme west side of anomaly 34. This, together with the fact that the rocks are of normal polarity and are overlain by sediments of Campanian age indicates that the basalts are of latest CNS age.

Sixty-six samples were selected for the experiment (12 each from Holes 417D, 418A and 543A, and 30 from Hole 807C). Figure 2 shows the results of a Thellier experiment on a representative example from Hole 807C on the Ontong Java Plateau. Most samples show a similarly rather ‘beautiful’ performance

during the Thellier test, despite their age. The samples usually have a maximum of two components: a low-temperature overprint, that can be removed by thermal demagnetization to a maximum of 250 °C (probably due to sample preparation¹³) and a well defined characteristic component. The susceptibility begins to increase around temperatures of 450 to 500 °C, so that in general only the temperature interval between ~200 and ~500 °C can be used for the palaeointensity calculation. During this interval, however, the sample does not alter, as indicated by constant χ , linear Arai plots and excellent pTRM reproducibility. Of the 66 samples, we were able to use 18 for palaeointensity calculations for Hole 807C, 11 each for Holes 417D and 418A, and 10 for Hole 543A. A success rate of more than 75% is exceptionally high, considering that other palaeointensity studies on submarine extrusives came to the conclusion that these rocks were more or less unsuitable for palaeointensity work (for example, refs 20–22). The glasses from Holes 417D and 418A show mean intensities of $12.0 \pm 3 \mu\text{T}$ and $15.5 \pm 8.8 \mu\text{T}$, respectively. The mean palaeointensity for 807C is $19.4 \pm 7.4 \mu\text{T}$ and that for 543A is $7.9 \pm 2.2 \mu\text{T}$. The glass shards used in this study are unoriented pieces from pillow rinds, so that in the absence of inclination data, it is necessary to translate the intensities into a quantity called the virtual axial dipole moment (VADM) in order to compare results globally. The VADM simply reduces the geomagnetic field to a rotationally symmetric dipole. Assuming a palaeolatitude of 17° N for 417D and 418A (ref. 23) we calculate VADM's of $2.8(\pm 0.7) \times 10^{22} \text{ Am}^2$ and $3.5(\pm 1.93) \times 10^{22} \text{ Am}^2$, respectively. With a VADM of $3.6(\pm 1.4) \times 10^{22} \text{ Am}^2$ for Hole 807C (assuming a palaeolatitude of 34° S, ref. 24) the results are remarkably similar for these almost contemporaneous but spatially distant drill holes. For Hole 543A, we calculated a palaeolatitude of 9° S, using the Upper Cretaceous pole of Van der Voo²⁵. This is consistent with an observed average inclination of -21° (ref. 26) and translates into a VADM of $2.0(\pm 0.6) \times 10^{22} \text{ Am}^2$ suggesting a low palaeo-field at the beginning and end of the CNS.

Cretaceous palaeointensities and the Geodynamo

In Fig. 3a we plot our palaeointensity results together with a slightly modified version of Prévot *et al.*'s palaeointensity compilation. Figure 3b shows a histogram of reversals against time as calculated from a combination of the geomagnetic polarity timescales of Cande and Kent¹⁸ (0–84 Myr) and Harland *et al.*⁶ (83–158 Myr).

Several conclusions can be drawn from the data in Fig. 3. (1) The intensity of the geomagnetic field in the CNS is low, measuring only ~45% of the intensity of today's field at the beginning, and decreasing to ~25% at the end. This extension of the 'Mesozoic dipole low' of Prévot *et al.* into the Cretaceous Superchron clearly does not support the strong fields during the CNS proposed by Larson and Olson⁴. (2) The intensities shortly before and after the ending of magnetic field reversals are essentially the same, thereby confirming earlier suggestions that there seems to be no link between changes in the reversal rate and changes in palaeointensity⁹. In addition it disproves the recently published interpretation that magnetization intensity variations of marine magnetic anomalies during the Cretaceous reflect changes in the palaeointensity of the magnetic field²⁷. (3) The absolute palaeointensity record is clearly still scanty. There are several large gaps, only one of which is the CNS. The lack of data during this period is therefore not a result of unknown peculiarities of this time interval but rather reflects the difficulty of obtaining reliable palaeointensity results in general.

Larson and Olson's⁴ strong-field model failed our test. This could mean several things; (1) there is no increase in convection in the outer core as a result of superplume volcanism at the beginning of the CNS, (2) stronger convection does not lead to a stronger dipole field, (3) there is no connection between the palaeointensity signal and changes in outer core convection, and (4) the samples cover too short a time window to average-out

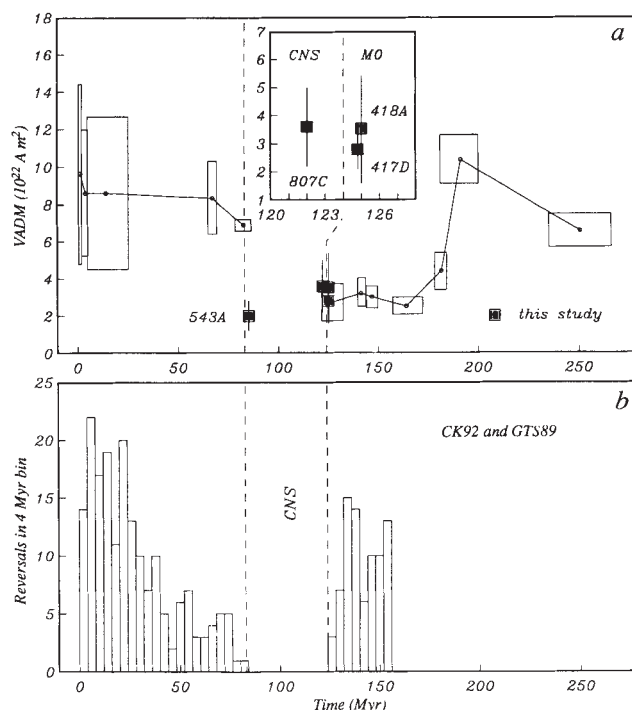


FIG. 3 Palaeointensity (virtual axial dipole moment, VADM) results together with compilation of Prévot *et al.*⁹ (a) and histogram of reversals against time (b). Data from cores from holes 417D and 418A are contemporaneous but data for core from hole 417D has been shifted slightly for better resolution. Intensities bracket the CNS and there seems to be no link between changes in the reversal rate and changes in palaeointensity.

secular variation of the magnetic field (SV). Possibilities (2) and (3) counter Olson and Hagee's²⁸ dipole model. Option (4) seems unlikely for the following reasons. Data from Holocene glasses show that secular variation of the geomagnetic field causes intensity variations of up to 3 times today's field over the last 5,000 yr¹³. The time necessary to average out SV is the subject of some debate, but most studies suggest that an interval of 10–20 kyr should be sufficient (for example, ref. 29). The drill cores investigated here have been sampled throughout the recovered basement. At Holes 417D and 418A, 253.2 and 389.7 m of basalt core were recovered, respectively. Together they comprise 7 distinct extrusive cycles (based on geochemical³⁰ and magnetic²³ evidence) which all have been sampled. Hole 807C penetrated 148.7 m of fresh basalt representing 5 volcanic subunits¹⁵ and 44 m of basalt (including 3 extrusive cycles²⁶) were sampled during the drilling of Hole 543A. Because of the differences in composition of these extrusive cycles an argument can be made (from a petrologist's point of view) that several hundreds to several thousands of years lie between these cycles³⁰. Together with the variation in the inclination data downcore^{23,26,31} this suggests that secular variation most definitely has been averaged out between Holes 417D, 418A and 807C. Although Natland *et al.*³² claim that the volcanics of Hole 543A bracket a large enough time slice to average out SV, a bias cannot definitely be ruled out for only three extrusive cycles.

Our data suggest that although both reversal rate and palaeointensity are signals that may conceivably originate in the outer core, they show very different signatures. While the reversal rate changes progressively over timescales of the order of 100 Myr, the average geomagnetic intensity (based on much more scanty data) appears to 'jump' between two values; intensities of the order of today's field and low field strengths during most of the Mesozoic era.

Recent evidence that reversals are linked to a decrease in intensity of the magnetic field^{18,33} belong in the realm of secular

variation and our results, along with the palaeointensity compilation of Prévot *et al.*⁹, suggest that the processes governing field reversals and long-term changes in palaeointensity are quite different.

As the oceans contain crust of late Jurassic to recent age, and, although not abundant, old glass can be found, deep drilling of the oceanic crust offers a great potential for future palaeointensity studies. Although glass has been recovered at many sites it is impossible to give an assessment of the availability of SBG at this stage. For the CNS we looked at cores from about

20 holes that supposedly had glass in them. In ten holes we found glass, but of these only four contained glass suitable for palaeointensity studies (that is, enough fresh material with strong remanent magnetization). Of course, the situation will improve as the rocks get younger but will get worse as we go back in time. Additional palaeointensity information for the Mesozoic and Cenozoic eras is necessary to shed more light on the possible connection between long-term intensity variations and CMB processes, and SBG is likely to be a rich source of such data. □

Received 2 August; accepted 22 October 1993.

1. Cox, A. V. *J. geophys. Res.* **73**, 3247–3260 (1968).
2. Loper, D. E. *Geophys. Res. Lett.* **19**, 25–28 (1992).
3. Merrill, R. T. & McFadden, P. L. *Science* **248**, 345–350 (1990).
4. Larson, R. L. & Olson, P. *Earth planet. Sci. Lett.* **107**, 437–447 (1991).
5. Loper, D. E. & McCartney, K. *Geophys. Res. Lett.* **13**, 1525–1528 (1986).
6. Harland, W. B. *et al.* *A Geologic Time Scale* (Cambridge Univ. Press, 1989).
7. Courtillot, V. & Besse, J. *Science* **237**, 1140–1147 (1987).
8. McFadden, P. L. & Merrill, R. T. *Physics Earth planet. Inter.* **43**, 22–33 (1986).
9. Prévot, M., Derder, M. E. M., McWilliams, M. & Thompson, J. *Earth planet. Sci. Lett.* **97**, 129–139 (1990).
10. Bol'shakov, A. S., Solodovnikov, G. M. & Vechfinsky, V. S. *Izvestiya Earth Physics* **14**, 904–910 (1978).
11. Sherwood, G., Shaw, J., Baer, G. & Mallik, S. B. *J. Geomagn. Geoelect.* **45**, 339–360 (1993).
12. Thellier, E. & Thellier, O. *Ann. Geophys.* **15**, 285–378 (1959).
13. Pick, T. & Tauxe, L. *J. geophys. Res.* **98**, 17949–17964 (1993).
14. Nagata, T., Arai, Y. & Momose, K. *J. geophys. Res.* **68**, 5277–5281 (1963).
15. Mahoney, J. J., Storey, M., Duncan, R. A., Spencer, K. J. & Pringle, M. *Proc. ODP Sci. Res.* **130**, 3–22 (1993).
16. Ozima, M., Kaneoka, I. & Yanagisawa, M. *Init. Rep. DSDP Leg 51–53*, 1127–1128 (1980).
17. Kent, D. V. & Gradstein, F. M. *Geol. Soc. Am. Bull.* **96**, 1419–1424 (1985).
18. Cande, S. C. & Kent, D. V. *J. geophys. Res.* **97**, 13917–13951 (1992).
19. Harland, W. B. *et al.* *A Geologic Time Scale* (Cambridge Univ. Press, 1982).
20. Bleil, U. & Petersen, N. *Nature* **301**, 384–388 (1983).

21. Grommé, S., Mankinen, E. A., Marshall, M. & Coe, R. S. *J. Geophys. Res.* **84**, 3553–3575 (1979).
22. Prévot, M., Mankinen, E. A., Grommé, S. & Lecaille, A. *J. geophys. Res.* **88**, 2316–2326 (1983).
23. Levi, S. *Init. Rep. DSDP Leg 51–53*, 1363–1378 (1980).
24. Tarduno, J. A. *et al.* *Science* **254**, 399–403 (1991).
25. Van der Voo, R. *Rev. geophys.* **28**, 167–206 (1990).
26. Wilson, D. S. *Init. Rep. DSDP Leg 78*, 583–592 (1984).
27. Sayanagi, K. & Tamaki, K. *Geophys. Res. Lett.* **12**, 2369–2372 (1992).
28. Olson, P. & Lee Hagee, V. *J. geophys. Res.* **95**, 4609–4620 (1990).
29. Opdyke, N. D. *Rev. Geophys. Space Phys.* **10**, 213–249 (1972).
30. Robinson, P. T., Flower, M. F. J., Swanson, D. A. & Staudigel, H. *Init. Rep. DSDP Leg 53*, 1535–1556 (1980).
31. Mayer, H. & Tarduno, J. A. *Proc. ODP Sci. Res.* **130**, 51–59 (1993).
32. Natland, J. H., Tarney, J., Marsh, N. G., Melson, W. G. & O'Hearn, T. *Init. Rep. DSDP Leg 78*, 393–400 (1984).
33. Gubbins, D. *Nature* **326**, 167–169 (1987).
34. Wohlfarth, E. P. *J. Appl. Phys.* **29**, 595–596 (1958).

ACKNOWLEDGEMENTS. We acknowledge the help of S. DiDonna in the laboratory, and the numerous conversations with C. Constable, J. Gee, M. Pringle and especially H. Staudigel who first suggested basaltic glass for palaeointensity studies. We thank R. Hamon and M. Perfit for supplying the 'BBQ' samples. The review of R. Coe greatly improved the manuscript. This study was supported by the US National Science Foundation, and we thank the Keck foundation for providing equipment. L.T. gratefully acknowledges partial support from the University of Utrecht.

LETTERS TO NATURE

Gravitational microlensing, quasar variability and missing matter

M. R. S. Hawkins

The Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

QUASARS have long been known to vary in magnitude¹, and it now appears that all quasars are to some extent variable². In a few extreme cases, quasar luminosities have varied over several magnitudes on a timescale of months; this behaviour is normally accompanied by other phenomena (such as radio emission or enhanced polarization) indicative of processes intrinsic to the quasars. Long-term luminosity variations, while less dramatic, are much more common, but their origin remains poorly understood. Here I investigate this longer-term behaviour for a sample of approximately 300 quasars at redshifts ranging from 1 to 3, whose optical magnitudes have been measured periodically for 17 years. I show that there is a positive correlation between the timescale of variability and the average luminosity, but little evidence for an increase in timescale due to redshift (and hence time dilation). These findings are inconsistent with any known variability mechanism intrinsic to quasars, but can be explained by gravitational lensing of the quasar images by compact substellar objects along the lines of sight. If this interpretation is correct, the population of lensing objects must have a density of at least 0.1 of the critical cosmological density.

Over the past 15 years or so a large-scale monitoring programme of quasars has been carried out using measures from the COSMOS high-speed microdensitometer at the Royal Observatory, Edinburgh, of a long sequence of wide-field photographs from the UK 1.2 m Schmidt telescope in Australia. The plates

were taken from 1975 to 1991, in ESO/SERC field 287 at right ascension 21 h 28 min, declination -45° . COSMOS then scanned each plate detecting and measuring around 200,000 stars and galaxies. Quasar candidates were selected solely on the basis of variability, with a minimum amplitude of 0.35 mag. This resulted in a sample of $\sim 1,000$ quasar candidates with light curves sampled every year for 17 years in an area of 19 square degrees. Details of the observational data and quasar selection procedure are given in refs 2–4. Various arguments were used, involving comparison with other samples in the field selected by ultraviolet excess and by objective prism spectra, to suggest that nearly all these candidates were indeed quasars. Follow-up spectroscopy has confirmed this, and reliable redshifts have now been assigned to ~ 300 quasars, which form the sample to be investigated for the nature of their variability. Most of the quasars have quite well defined variability characteristics, with a timescale between minimum and maximum of typically ~ 5 yr. The variation appears quasi-sinusoidal with little evidence of flares or eclipses. Figure 1a and b shows two typical light curves.

In recent studies of quasar variability two possible correlations have been discussed. First, there are a number of general theoretical arguments that suggest that a positive correlation between luminosity and timescale should exist. This question has been addressed in several observational studies^{2,5} but it seems fair to say that there is no conclusive evidence to support this hypothesis. Second, if quasars have some intrinsic preferred timescale of variation then under the most general considerations time dilation will increase the timescale by a factor of $(1+z)$, producing a correlation between redshift (z) and timescale^{2,5}. This correlation has not been definitely observed to date.

To investigate the causes of variability in the sample of quasars it is necessary to have some way of characterizing the variability timescale of subsets of the main sample. The time-varying autocorrelation function (ACF) provides a convenient way of

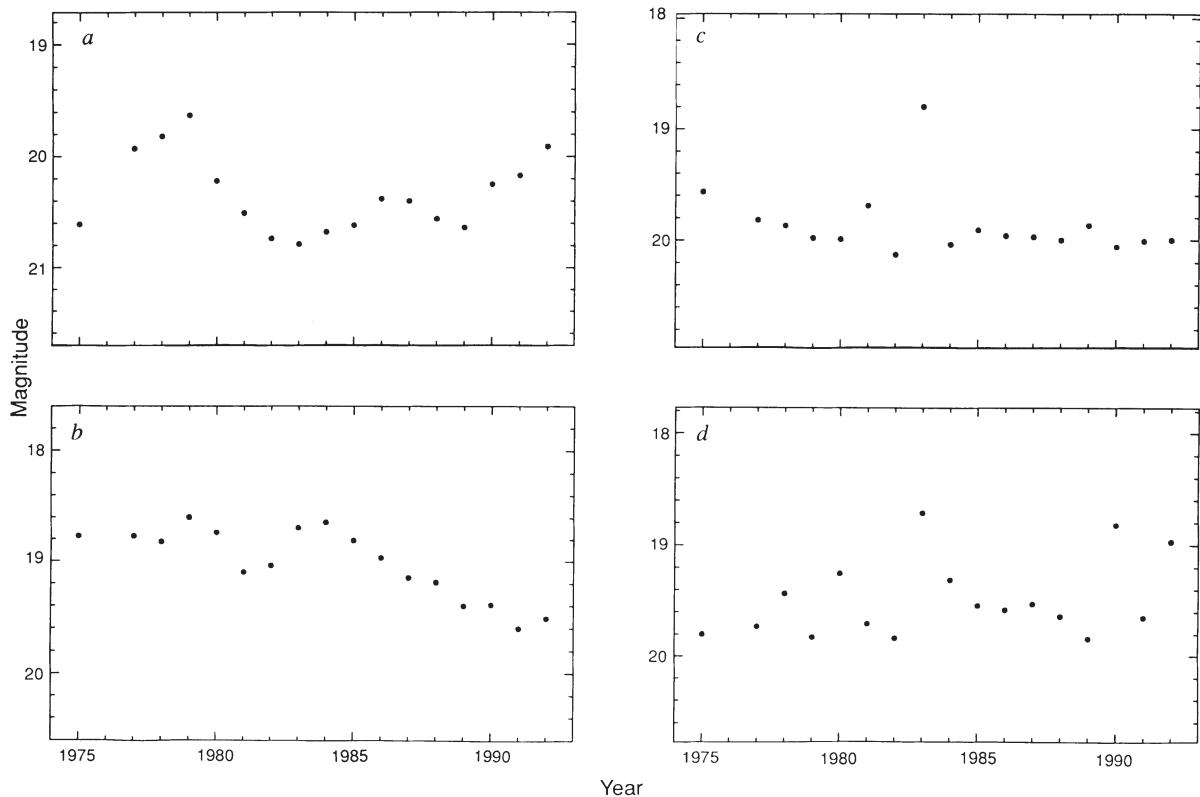


FIG. 1 Light curves for quasars from the sample. Magnitude errors are ~ 0.05 mag (twice the size of the dots). *a* and *b*, Two typical quasars

with redshift $z \approx 2$. *c* and *d*, Two low-redshift quasars with $z \approx 0.2$.

doing this. It is defined by

$$A(t) = \frac{\sum m(s) \cdot m(s+t)}{\sum m(s) \cdot m(s)}$$

where $m(s)$ is the departure from mean magnitude at the epoch of observation s . Thus $A(t)$ is the amplitude of the ACF at time t . This function effectively describes the time over which magnitudes are correlated, and hence the timescale of variation. In order to investigate these correlations in the present data, the sample was sorted by ('binned') luminosity and redshift, assuming that the current value of the Hubble constant $H_0 = 50$ and the deceleration parameter $q_0 = 0.5$. Each ACF was then fitted with a curve of the form

$$y = b(\exp(-ax) - 1) + 1$$

where a and b are free parameters, and the timescale is given by $t = -\ln(0.5)/a$. The results are shown in Fig. 2. The two low-luminosity bins (Fig. 2*a, b*) have timescale 1.4 and 2.3 yr for high (Fig. 2*a*) and low (Fig. 2*b*) redshift respectively. Equivalent figures for high luminosity (Fig. 2*c, d*) are 3.5 and 5.2. Errors on these timescales were evaluated by splitting the bins into two halves, and are ± 0.2 yr or less. For each redshift bin there is an increase in timescale by a factor 2.4 for a 1 mag increase in luminosity. But for an increase in mean redshift from 1.5 to 2.5, the timescale actually decreases in both luminosity bins by a factor of ~ 0.7 . This remarkable observation suggests that the variation cannot be intrinsic to the quasars. The significance of the shape of the ACF may be tested by randomly permuting the epochs, and then calculating the ACF as before. In all cases the function is flat and lies close to zero, as would be expected.

There have been few attempts to date to describe quasar variability from a theoretical standpoint. The only timescale which emerges naturally from quasar structure theories (see ref. 6 for a review) is the few weeks associated with light travel time across

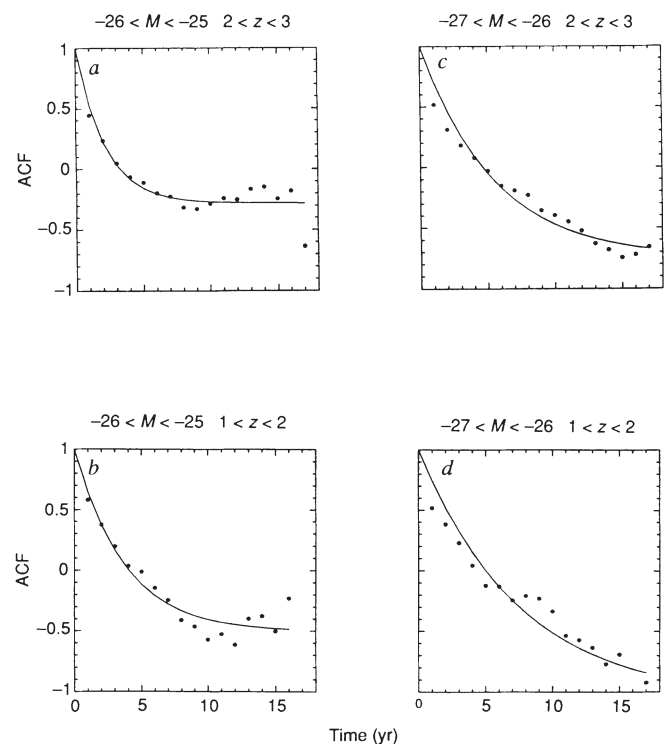


FIG. 2 Time-varying autocorrelation functions (ACF) for the quasar sample, sorted by luminosity (magnitude M) and redshift (z). Each ACF is fitted by an exponential curve defined in the text.

the central black hole and accretion disc. This timescale is observed in optically violently variable (OVV) quasars, but not in the average quasar light curve. Possible mechanisms for longer-term variability include 'weather systems' in the broad line region, and instabilities in the accretion disc analogous to those seen in dwarf novae, but none of these processes produce variations similar to those actually observed⁶.

The idea that some quasar variation may be due to microlensing by stars in intervening galaxies has been examined extensively (for example, refs 7–10). (Model light curves, which can in principle be compared to observations, are calculated for various configurations of lens and source.) The attractive aspect of this idea is that the observed variation is subject to time dilation at the redshift of the intervening galaxy, and not the quasar. It also produces model light curves which bear a strong resemblance to those actually observed. For example the light curves in Fig. 1 may be compared with the large source sizes of Fig. 2 in ref. 9. For redshifts $> \sim 0.5$ the observed angular diameter of the quasar disc will only become slightly smaller with redshift, changing by a factor of 0.8 from $z=1.5$ to $z=2.5$. This is close to the change in timescale in Fig. 2 (this work), and consistent with the hypothesis that the timescale of variation is related to the time taken for a lensing object to cross the line of sight.

The distance over which the microlensing object significantly influences the light from the source is known as the Einstein radius, which varies as the square root of the lens mass¹¹. It is the crossing time of the associated Einstein ring that determines the timescale of variation of the source. The correlation between luminosity and timescale is a consequence of the relation between quasar disc size and luminosity; a larger disc results in a longer crossing time of the Einstein ring of the lensing object, and hence a longer timescale of variation. It is not possible to make a quantitative comparison with the observations without a specific model for the temperature distribution across the disc of a quasar of given luminosity. It is, however, clear that the quasar disc must be comparable in size to the Einstein ring of the lens for it to affect the crossing time and give the observed correlation.

If microlensing is the cause of variability, then statistically there should be symmetry between the rising and falling points

of the light curves. Figure 3 shows histograms of magnitude increases and decreases for the present quasar sample, and it can be seen that the two distributions are indistinguishable. Clearly, at sufficiently low redshift, quasars will not be lensed. Figure 1c and d shows two low redshift variable quasars from the sample. The nature of their variability is completely different from that of most of the quasars in the sample (compare Fig. 1a and b); either we are witnessing intrinsic short-lived flares, or occasional isolated microlensing events. Both these possibilities are consistent with the hypothesis that in general quasar variability is caused by microlensing. Further support for this interpretation comes from the autocorrelation function for the lowest redshift quasars in the sample ($z < 0.3$), which is flat and close to zero as would be expected from random short-lived outbursts.

Press and Gunn¹² have shown that for a Universe in which every line of sight is microlensed, $\Omega_\lambda \approx 1$, where Ω_λ is the contribution to the cosmological density parameter in the lensing material. Given a set of lensing light curves, it is not simple to measure the value of Ω_λ , but from model light curves available in the literature it is possible to make some preliminary estimates. All simulations published to date assume an optical depth of lenses of at least 0.1 (refs 8–10), which corresponds to values of Ω_λ in the range 0.5–1.0 (ref 7). In this case, for point sources, amplitudes are predicted of ~ 3 mag (refs 8, 9) whereas for extended sources this reduces to ~ 0.5 –1.0 mag and the light curves become much smoother. This amplitude is still larger than that observed (~ 0.5 mag), and so is likely to correspond to a smaller value of Ω_λ . Further simulations using smaller optical depths should help to give a more precise figure.

The timescale of variability for a microlensed quasar is generally taken to be the crossing time of the Einstein ring (refs 11, 12). Thus given the relative transverse velocity and timescale, one can calculate the diameter of the Einstein ring and hence the mass of the lens. For a velocity of 600 km s^{-1} and a timescale of 10 yr, the lens mass is ~ 0.05 solar masses, ($M_\odot$). If it is the case, as argued above, that the quasar disc is of similar size to the impact parameter of the lens, the quasar diameter is ~ 0.01 pc or 10 light d. This corresponds to the smallest timescales observed for OVV variability. One consequence of this¹³ is that if contin-

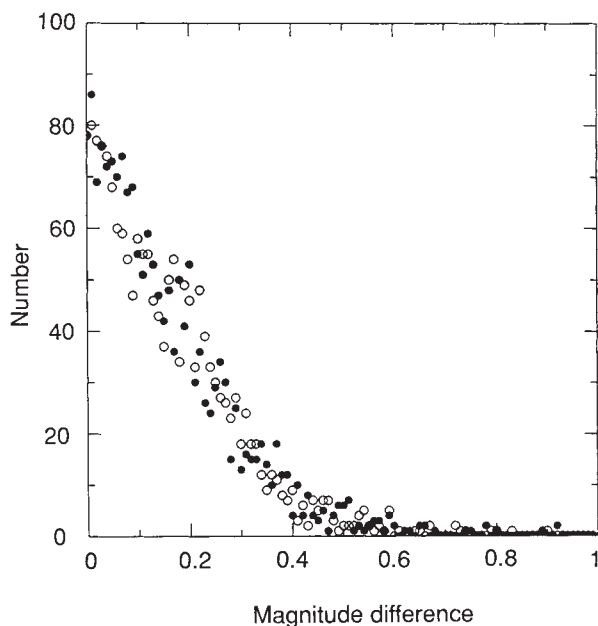


FIG. 3 Histograms of magnitude increases (open circles) and decreases (filled circles) from year to year from the quasar sample.

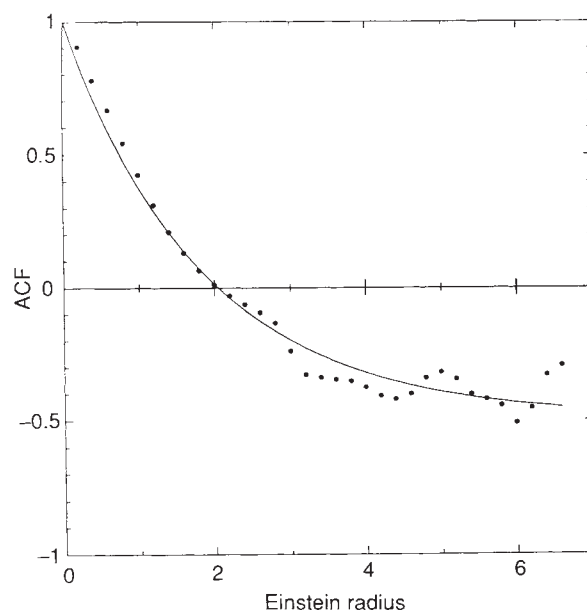


FIG. 4 Autocorrelation function (ACF) for simulated microlensing light curves from ref. 14. Lenses are distributed in a sheet with optical depth of 0.5, and source size 0.1 Einstein radii.

uum and broad line emission come from very different sized regions (as predicted by current quasar models), there should be changes in the emission line strength as a quasar is microlensed. The data available now can do no more than limit Ω_λ to <1 (ref. 13) but clearly this is a possible way to test the microlensing hypothesis.

A more direct comparison between observation and theory can be made using a long run of simulated data recently published by Lewis *et al.*¹⁴. Lenses are distributed in a sheet with optical depth of 0.5 and source size 0.1 Einstein radii. Figure 4 shows the ACF measured from Fig. 3b of ref 14, together with the fitted function as for Fig. 2 (this work). The timescale is 1.3 Einstein radii which, when compared with the corresponding timescales from Fig. 2, gives lens masses of $\sim 2 \times 10^{-3} M_\odot$. Inspection of the simulated data suggests that this lower mass is the result of groups of microlensing events dominating the ACF, rather than single transits. It seems clear that convincing measurements of the lensing masses can only be made using

simulations tailored to the observations, and this will form the subject of a future paper. □

Received 14 July; accepted 22 September 1993.

1. Penston, M. V. & Cannon, R. D. Bulletin No. 159 (Royal Observatory, Edinburgh, 1970).
2. Hawkins, M. R. S. & Veron, P. *Mon. Not. R. astr. Soc.* **260**, 202–208 (1993).
3. Hawkins, M. R. S. *Mon. Not. R. astr. Soc.* **202**, 571–585 (1983).
4. Hawkins, M. R. S. *Mon. Not. R. astr. Soc.* **219**, 417–426 (1986).
5. Hook, I. M., McMahon, R. G., Boyle, B. J. & Irwin, M. J. in *The Space Distribution of Quasars ASP Conf. Ser. Vol. 21* (ed. Crampton, D.) 55 (Astronomical Society of the Pacific, San Francisco, 1991).
6. Rees, M. J. *A. Rev. Astr. Astrophys.* **22**, 471–506 (1984).
7. Peacock, J. A. *Mon. Not. R. astr. Soc.* **223**, 113–128 (1986).
8. Kayser, R., Refsdal, S. & Stabell, R. *Astr. Astrophys.* **166**, 36–52 (1986).
9. Schneider, P. & Weiss, A. *Astr. Astrophys.* **171**, 49–65 (1987).
10. Refsdal, S. & Stabell, R. *Astr. Astrophys.* **250**, 62–66 (1991).
11. Blandford, R. D. & Narayan, R. A. *Rev. Astr. Astrophys.* **30**, 311–358 (1992).
12. Press, W. H. & Gunn, J. E. *Astrophys. J.* **185**, 397–412 (1973).
13. Canizares, C. R. *Astrophys. J.* **263**, 508–517 (1982).
14. Lewis, G. F., Miralda-Escudé, J., Richardson, D. C. & Wambsganss, J. *Mon. Not. R. astr. Soc.* **261**, 647–656 (1993).

ACKNOWLEDGEMENTS. I thank R. Webster for pointing out the possibilities of microlensing, and J. Peacock and L. Miller for suggestions.

Imaging high-energy astrophysical sources using Earth occultation

S. N. Zhang^{*†}, G. J. Fishman^{*}, B. A. Harmon^{*} & W. S. Paciesas^{*‡}

^{*} NASA Marshall Space Flight Center, Code ES66, Huntsville, Alabama 35812, USA

[†] Universities Space Research Association, Huntsville, Alabama 35806, USA

[‡] University of Alabama in Huntsville, Huntsville, Alabama 35899, USA

HIGH-ENERGY astrophysical sources can be difficult to image. Photons with energies above ~ 5 keV are hard to focus, so experiments usually employ coded masks^{1–4} or moving collimators^{5–9} to modulate the flux received by the detectors; the resulting signals are then deconvolved to form the images. Here we demonstrate a new approach which makes use of the large-area, non-collimated detectors of the Burst and Transient Source Experiment on the Compton Gamma-Ray Observatory. As the spacecraft moves in its orbit, the Earth itself acts as a stable occulting disk. Changes in the measured signal during a single occultation correspond to the integrated intensity of sources positioned along the arc described by the Earth's edge (limb). The low-altitude, moderately inclined orbit of the spacecraft ensures that the angle at which the limb traverses a source region varies between occultations, and thus data from a series of occultations can be transformed into an image. This imaging process is conceptually and mathematically similar to those used in fan-beam aperture-synthesis radio-astronomy¹⁰ and medical computer-assisted tomography¹¹, and holds great promise for all-sky imaging with relatively simple (and hence inexpensive) detectors.

In visible-wavelength optics, transform imaging has been described for at least the past 30 years^{12,13}, although it has not seen widespread applications. For high-energy astronomy, transform-imaging processes are reviewed by Caroli *et al.*⁷ and by Lei *et al.*¹⁴ The coded-aperture imager produces a spatially modulated signal, whereas the rotation-modulation imager produces a temporal modulation. The modulated signal from the detector is later deconvolved to produce an image. A variant of the rotation-modulation imager is a scanning-modulation imager, in which scans of the imaged region are made along different scan angles. Experiments on previous spacecraft such as Uhuru, SAS-3, Ariel-5 and HEAO-1 have used 'lines of position'

perpendicular to the scan directions of a collimator to locate point sources. They could not, however, obtain a true transformed image of a region (as is obtained here) because of the limited number of scan directions available and/or the long time between scans. These previous techniques also require a pointed or scanning mask with a size comparable to that of the detector. The field-of-view is limited by the mask size, and the angular resolution is limited by the distance between the mask and detector.

Our technique uses the Earth's limb as a well defined, stable occulting disk. We call this process occultation-transform imaging. The features of the three imaging techniques are compared in Table 1. This technique is demonstrated using the large-area detectors of the Burst and Transient Source Experiment¹⁵ (BATSE) on the Compton Gamma-Ray Observatory (GRO). The typical resolution (point-spread function) for the BATSE occultation-imaging system is $\sim 0.5^\circ$ and the positional accuracy is $\sim 0.1^\circ$. A significant component of the positional error results from the limited time resolution of the BATSE data. Another component, of comparable magnitude, results from the 'fuzziness' of the Earth's limb (arising from the atmosphere) as sources are occulted. In principle, this latter source of image degradation can be reduced as the grazing atmospheric transmission can be accurately modelled as a function of energy.

At a typical GRO altitude of 400 km, the Earth's limb for occultation is about 2,000 km away and lies about 70 km above the Earth's surface for γ -rays of energy 100 keV and approximately 65 km above the Earth at 1 MeV. The limb is about 108° from the zenith. At any given instant, the Earth subtends 34% of the total solid angle, as seen from the spacecraft. Throughout one orbit, sources in approximately 70% of the sky can produce two useful occultations per orbit (emergence/immersion or rise/set). As the GRO orbit precesses (with a period of 53 days), both of these occultation angles change in a continuous manner. For a low-altitude spacecraft with a moderate inclination, all sky regions eventually undergo occultation as the orbit precesses. The GRO orbit is inclined 28.5° with respect to the Earth; occultations are obtained for sources $>47^\circ$ from the orbital poles, but the available scanning time for sources near the celestial poles is reduced by about 30% relative to that of other sky regions.

Observations of sources by occultation were planned by the BATSE group before launch¹⁵. The original BATSE occultation method was one-dimensional in nature, or combined two or more one-dimensional scans across the source region to isolate and locate the source under study. It can be used only for well separated, relatively strong sources.

Our imaging process is an extension of the one-dimensional technique. When the limb of the Earth (an arc) sweeps through

TABLE 1 Comparison of hard X-ray/ γ -ray imaging techniques for astronomy

	Coded-aperture transform imaging	Rotation- (or scanning-) modulation imaging	Occultation-transform imaging
Experiment-spacecraft	SIGMA-GRANAT	WATCH-GRANAT	BATSE-GRO
Field-of-view	$\sim 10^\circ \times 10^\circ$	$\sim 120^\circ$ FWHM (per detector)	Whole-sky (8 detectors)
Mask	50% opaque 2-D array	50% opaque slat mask	No mask (Earth limb)
Modulation produced by:	Stationary mask	Rotating (or scanning) mask, grid or slats	Spacecraft orbital motion; orbit precession
Sensitivity at edge of field	Low	Low	High
Position sensitive detector:	Required	Not required	Not required
Energy limitations:	Mask transmission, detector response	Mask transmission, detector response	Detector response
Inherent limitations on angular resolution	Detector spatial resolution, mask element size, mask-detector separation	Slat size, slat-detector separation	Earth limb location and model
Inherent limitations on temporal resolution	No limitation	Mask rotation (scan) period	Orbital period
References (Typ.)	1-4, 7, 14	5-9, 14	This paper

a chosen region of the sky, the detectors record the counting rate as a function of time. The occultation of a source is typically detected as a small ($\lesssim 1\%$) but rapid change in the counting rate in a detector. This occurs over a time of about $8/\cos \beta$ seconds, where β is the angle of the source with respect to the orbital plane. The location of the spacecraft is well known and thus the location and orientation of the limb projected on the sky can be determined accurately. The counting rate produced as the limb traverses a source region is differentiated. This becomes a signal indicative of the intensity of the sources integrated along the arc. This signal, representing a continuous series of arcs at different projection angles, is used as the input to the Radon transformation^{11,16,17}, which inverts the data to produce an image. This resembles computer-assisted tomographic imaging in that both processes require a large set of input source angles to reconstruct a reasonable image. For the present work, we have used the maximum-entropy software package of the Donner algorithm for Radon transform imaging¹⁶. Numerous images have been produced with fields ranging from 5° to 40° on a side. For larger regions, the Radon transformation is not well behaved and the images are highly distorted.

The technique is demonstrated in a series of images in two sky regions. The first region is a wide ($40^\circ \times 40^\circ$) field which includes a strong, stable source (the Crab Nebula) and the intense, transient hard X-ray source GRO J0422+32 discovered by the BATSE experiment in August 1992^{18,19}. Three images are shown (Fig. 1a-c) of the same region, at different times. Figure 1a was made during the rising part of the outburst (intensity ~ 0.05 Crab); Fig. 1b was near the peak of the outburst (~ 3 Crab); and Fig. 1c (~ 1 Crab) was made during the declining phase. The offset of the image contours from the true source locations (shown by crosses) results from the distortion of the very wide field of the image. This offset is removed as the field is reduced. Images of strong sources have been positioned to within 0.1° in narrower-field (5°) images. Elongation of the contours around the point sources is a result of incomplete transformation, which stretches the contours along the direction of the Earth's limb.

The second region is a $20^\circ \times 20^\circ$ field near the Galactic Centre (Fig. 2). Four images are shown in this figure, taken at different times. The extreme variability of the hard X-ray sources in this region is evident and has been noted by other experiments with imaging capabilities²⁰⁻²⁸. From these images, only two sources in Fig. 2c and 2d, namely GX354-00 and 1E1740.7-2942, were unambiguously isolated and identified. GX1+4, shown in Fig. 1a, was also shown to be detected from other images produced with slightly different fields in the same period. This is in good

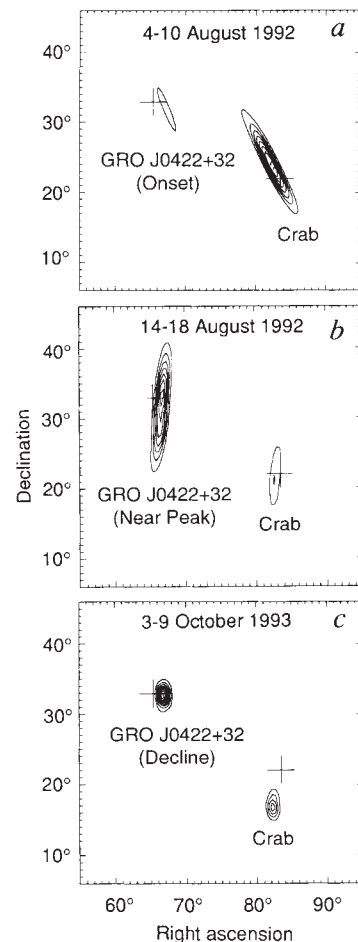


FIG. 1 a-c, The sky region containing the Crab Nebula and the intense, transient source GRO J0422+32 (Nova Persei 1992) at three different times during its outburst. These images have been produced by the method described in the text. A nonlinear, maximum-entropy processing of the images results in image contours that are not related to the absolute flux values, but instead indicate the relative intensity of sources within the field. The wide field of these images causes distortion such that the image contours are offset from the actual source locations (crosses). Narrow-field images of regions of strong sources yield contours that are within 0.1° of the actual location. The energy range for these images is 50 to 100 keV. The image size is $40^\circ \times 40^\circ$.

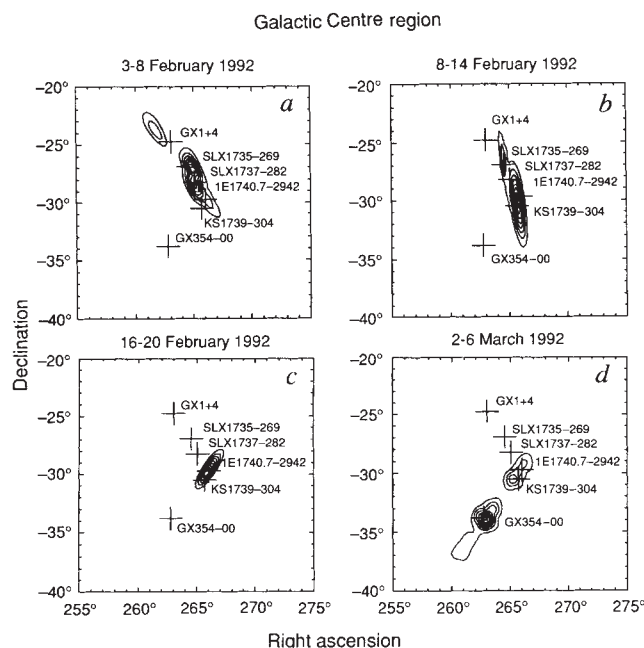


FIG. 2 a–d, Images of the Galactic Centre region ($20^\circ \times 20^\circ$) at four different times, showing at least four variable sources. The intense state of GX354-00 in Fig. 2d is consistent with observations of the SIGMA experiment on the GRANAT spacecraft. The variability of other sources has not been observed independently during these times. The crosses show the actual location of sources observed previously. As in Fig. 1, the source locations are offset because of the wide field of these images and because there are other sources present in the field. Elongation of image contours in Figs 1 and 2 arises from incomplete or unfavourable occultation angles during the observations.

agreement with SIGMA/GRANAT observations in the similar periods²⁸. The complex around the centre of the field shown in Fig. 2a and 2b agrees qualitatively with the significant detections of SLX1735–269 and 1E1740.7–2942 and the marginal detections of SLX1737–282 and KS1739–304 by SIGMA/

GRANAT²⁸. The present imaging system does not allow them to be separated unambiguously.

There are several instrumental considerations that make this technique particularly attractive for space-borne applications. There is no need for a large, massive mask and its support structure. Similarly, there are only minimal pointing requirements: the occulting disk (the Earth's limb) is stable, and it is sufficiently dense that it can be used effectively for higher-energy γ -rays. The greatest advantage of this technique over a coded-mask imager is that a position-sensitive detector is not required. An advantage over a rotation-modulation collimator is that there are no moving parts in the system, which have caused reliability problems in previous space-borne experiments. Other transform techniques have greatly reduced sensitivity when the source region is near the edge of the field. With occultation-transform imaging, there is no mask to restrict the field-of-view or reduce the off-axis sensitivity. The octahedral geometry of the BATSE detectors assures that high sensitivity can be achieved for source regions over the entire sky. Finally, for any imaging system utilizing a mask, the orientation of the mask with respect to the celestial sphere must be determined at all times. With the present technique, only the location of the spacecraft with respect to the Earth's limb need be determined to the required accuracy.

Many transient and highly variable sources, both galactic and extragalactic, vary on timescales observable by this imaging system—several days to years. Narrower-field instruments, of necessity, have a low observational duty cycle and thus have limited monitoring capability for most regions of the sky. The typical sensitivity achieved thus far is about 50 mCrab for a 1-week observation, as determined by observations of GRO J0422+32 during its declining phase.

Data analysis efforts are underway to process the large data set obtained by BATSE since launch (April 1991) and to image regions over the entire sky. Detection sensitivity, image quality and resolution are expected to improve beyond the preliminary images shown here as the background systematics become better known and as more-advanced image processing techniques are used with the data. As the BATSE detectors were designed for the study of γ -ray bursts, they are not optimized for this imaging application. Future high-energy astronomy experiments utilizing this technique can be improved greatly by using low background detectors and by collimating the detectors to view only a field around the horizon. This in turn, could permit the use of a simple, passive, gravity-gradient-stabilized spacecraft. □

Received 2 April; accepted 9 September 1993.

1. Fenimore, E. E. & Cannon, T. M. *Appl. Opt.* **17**, 337–347 (1978).
2. Skinner, G. K. *Nucl. Instrum. Meth.* **221**, 33–40 (1984).
3. Kohman, T. P. *Rev. Scient. Instrum.* **60**(11), 3396–3409 (1989).
4. Skinner, G. K. *Scient. Am.* **259**(2), 84–89 (August 1988).
5. Durouchoux, P., Hudson, H., Hurford, G., Hurley, K., Matteson, J. & Orsai, E. *Astr. & Astrophys.* **120**, 150–155 (1983).
6. Mertz, L. N., Nakano, G. H. & Kilner, J. R. *J. Opt. Soc. Am. A*, **3**, 2167–2170 (1986).
7. Caroli, E., Stephen, J. B., DiCocco, G., Natalucci, L. & Spizzichini, G. *Space Sci. Rev.* **45**, 349–403 (1987).
8. Willmore, A. P. *Mon. Not. R. Astr. Soc.* **147**, 387–403 (1970).
9. Cruise, A. M. *Mon. Not. R. Astr. Soc.* **170**, 305–312 (1975).
10. Bracewell, R. N. *Aust. J. Phys.* **9**, 198–217 (1956).
11. Gullberg, G. T. & Tsui, B. M. W. *Information Processing in Medical Imaging* (eds deGraaf, C. N. & Viergeren, M. A.) 181–189 (Plenum Press, New York, 1989).
12. Mertz, L. *Transformations in Optics* (Wiley, New York, 1965).
13. Mertz, L. in *Proc. EUV, X-ray and Gamma Ray Instrumentation for Astronomy and Atomic Physics*, SPIE **1159**, 14–17 (1989).
14. Lei, F., Fraser-Mitchell, J. & Yearworth, M. *Expl. Astr.* **1**, 285–303 (1991).
15. Fishman, G. J. et al. in *Proc. Gamma-Ray Observatory Science Workshop* (ed. Johnson, W. N.) 39–50 (NASA/GSFC, Greenbelt, MD, 1989).

16. Huesman, R. H. et al. *Donner Algorithms for Reconstruction Tomography*. LBL Publ. **214**, 42–44 (University of California, Berkeley, 1977).
17. Deans, S. R. *The Radon Transform and Some of Its Applications* (Wiley, New York, 1983).
18. Harmon, B. A. et al. *IAU Circ. No.* 5781 (1993).
19. Harmon, B. A. et al. in *AIP Conf. Proc.* **280**, 314–318 *Compton Gamma Ray Observatory* (eds Friedlander, M., Gehrels, N. & Macomb, D.) (1993).
20. Skinner, G. K. et al. *Nature* **330**, 544–547 (1987).
21. Claret, A., Goldwurm, A., Cordier, A. & Paul, J. *Astrophys. J.* (in the press).
22. Cook, W. R. et al. *Astrophys. J.* **372**, L75–L81 (1991).
23. Sunyaev, R. et al. *Astr. Astrophys.* **247**, L29–L32 (1991).
24. Sinner, G. *Astr. Astrophys. Suppl. Ser.* **97**, 149–153 (1993).
25. Grindlay, J. E., Covault, C. E. & Manandhar, R. P. *Astr. Astrophys. Suppl. Ser.* **97**, 155–158 (1993).
26. Bazzano, L. et al. *Astr. Astrophys. Suppl. Ser.* **97**, 169–171 (1993).
27. Cordier, B. et al. *Astr. Astrophys. Suppl. Ser.* **97**, 177–180 (1993).
28. Churazov, E. et al. *Astr. Astrophys. Suppl. Ser.* **97**, 173–176 (1993).
29. Paciesas, W. S. et al. in *AIP Conf. Proc.* **280**, 473–477 *Compton Gamma Ray Observatory* (eds Friedlander, M., Gehrels, N. & Macomb, D.) (1993).

ACKNOWLEDGEMENTS. We are grateful for the help of the BATSE operations and data analysis teams. We have benefited from discussions with M. Finger, B. Rubin, C. Meegan and R. Wilson. We thank R. H. Huesman for providing the Donner algorithms.

Stabilization of a metastable crystalline phase by twinning

R. J. Davey^{*†}, S. J. Magl^{nn*}, S. J. Andrews[†],
A. M. Buckley[†], D. Cottler[†], P. Dempsey[†],
J. E. Rout[†], D. R. Stanley[†] & A. Taylor[†]

^{*} Zeneca plc and [†] ICI Chemicals and Polymers, R & T Department, The Heath, Runcorn, Cheshire WA7 4QD, UK

[†] Present address: Department of Chemical Engineering, UMIST, PO Box 88, Manchester M60 1QD, UK

POLYMORPHISM of crystal structures—alternative crystal structures resulting from variations in ionic and molecular packing¹ and conformation²—is important in a number of fields in solid-state chemistry, including biomineralization³ (in the mineral phases of calcium carbonate³, for example), polymer science (polypropylene⁴, for example), explosives (ammonium nitrate, lead azide⁵), nonlinear optical materials⁶, ceramics and catalysis (zeolites and metal oxides⁷). In the pharmaceutical and fine-chemicals industries, polymorphism is central to both production process design and product activity⁸. Here we show that crystal twinning can stabilize a crystal polymorph that is otherwise not the most stable form. We use electron microscopy, X-ray diffraction and Raman spectroscopy to show that the apparently indefinite persistence of a metastable polymorph of terephthalic acid can be explained on this basis. If twinning can be engineered, it may therefore provide a means for stabilizing crystal phases with useful physical properties.

The crystallography of terephthalic acid (TA; $C_6H_4(COOH)_2$) was first determined by Bailey and Brown⁹ and later discussed by Leiserowitz and Berkovitch-Yellin¹⁰. Two polymorphs are known (forms I and II) each with triclinic unit cells, space group $P\bar{1}$. In both, molecules are linked through hydrogen-bonded carboxyl dimers into infinite chains. These chains are held in two-dimensional sheets by van der Waals and $-C-H\cdots O-$ interactions. The essential difference in the structures lies in the juxtaposition of the sheets. In form I the carboxyl dimers of one sheet lie directly above those in adjacent sheets, whereas in form II the dimers of one sheet lie over the

phenyl rings of adjacent layers. At the outset of this study the relative stability of the forms had not been clearly established. Bailey and Brown⁹ assumed that form I was more stable because it has the higher density and because form II 'tended to become rarer on storage'. Gerasimov *et al.*¹¹ continued this assumption in their Raman spectroscopic study of TA, whereas Saska and Myerson¹², using powder X-ray diffraction, were unable to find any evidence of form II.

In the experiments reported here, pure TA (99.9%, ICI Chemicals and Polymers Ltd) was dissolved in water and acetic acid in an autoclave at 250 °C and 4.5×10^6 Pa. Crystals formed on cooling such solutions were recovered at room temperature and pressure. The phases and morphologies of these, together with as-supplied commercial TA, were characterized by optical and Raman hot-stage microscopy, scanning electron microscopy, Weissenberg single-crystal X-ray diffraction and Laue single-crystal X-ray diffraction (using the synchrotron radiation, SR, source, at SERC, Daresbury, UK, and a temperature-controlled furnace¹³).

We found crystals to be in the size range 50–100 μ m. Form I crystals could be prepared only in the presence of low levels (5%) of added *p*-toluic acid and were always found to be *c*-axis needles twinned with common directions $[1\bar{1}0]$ and $[001]$; one of these is the same in both twins, the other exactly reversed. Because these directions are separated by an angle of 77.83°, the two lattices in contact across the twin plane are incommensurate. This suggests that defects are involved, perhaps associated with the incorporation of *p*-toluic acid and subsequent surface termination of the hydrogen-bonded chains during crystal growth. One possibility might be that twinning results from an alternative hydrogen-bonding pattern across the twin plane, promoted by the need for carboxylate groups in the next terephthalic acid layer to avoid surface methyl groups as the crystal grows. The consequence of this process is that the chain direction across the twin boundary is conserved but the molecular planes are rotated through 37.06°. This is shown in Fig. 1. Form II crystals grew either as single hexagonal *c*-axis needles or rhombs with dominant (100) facets. Commercial TA from a number of sources were examined and all found to be form I. A typical series of images from the scanning electron microscope (SEM) is shown in Fig. 2. These ovoid particles are interesting in that externally

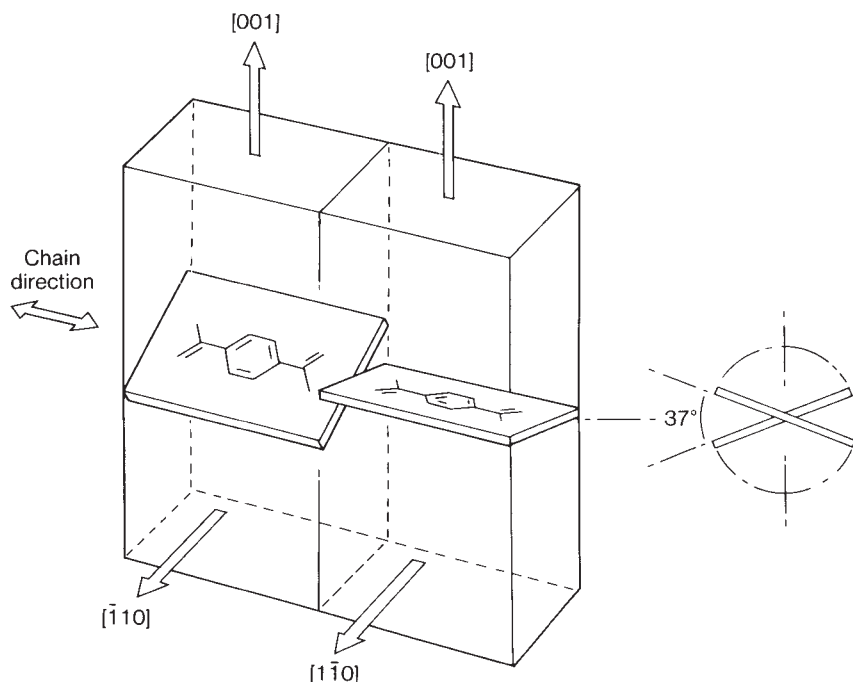


FIG. 1 Schematic relationship between form I twins of terephthalic acid (TA)

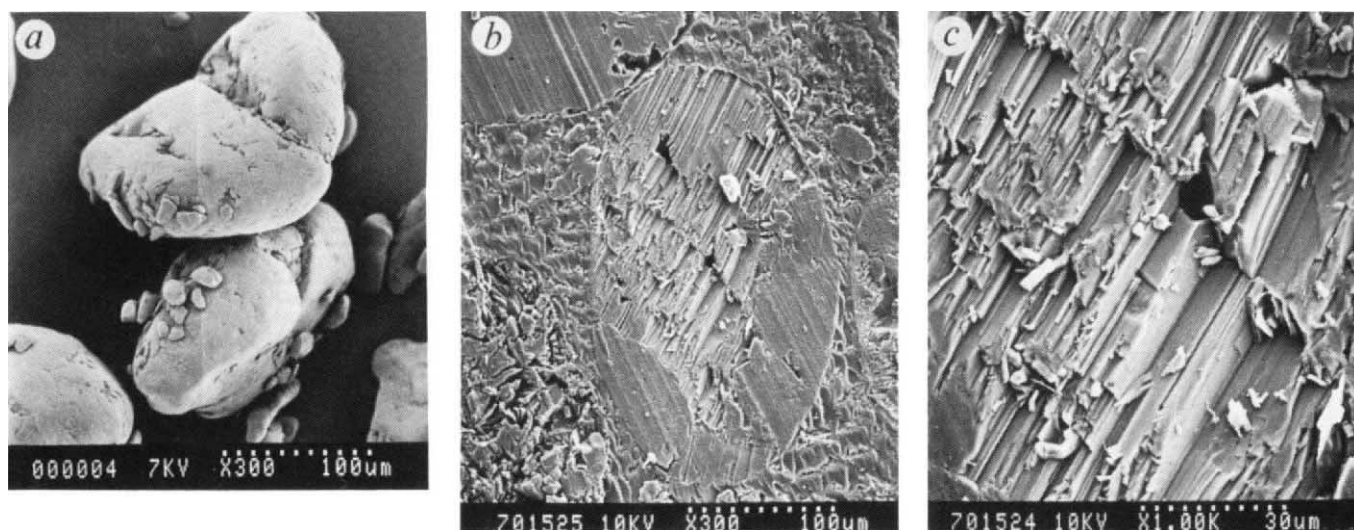
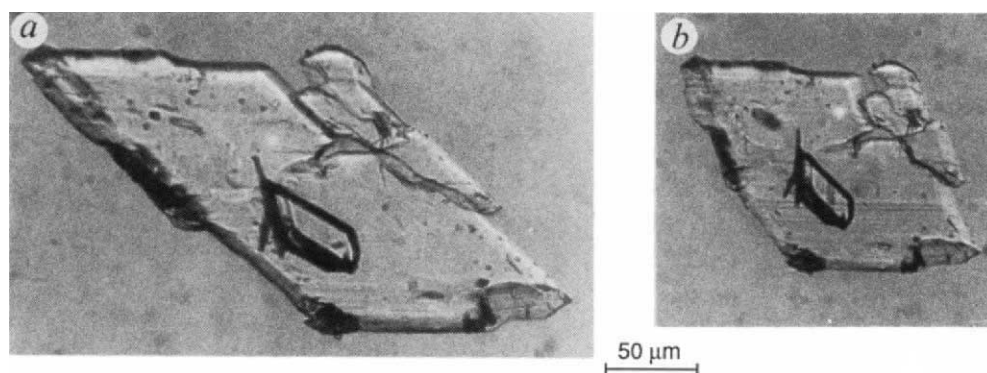


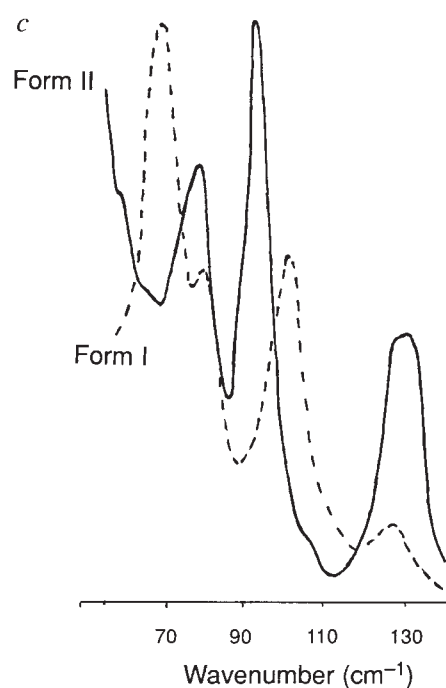
FIG. 2 The microstructure of commercial TA. Note that the dotted scale bars in a, b and c are 100, 100 and 30 μm respectively.

FIG. 3 Changes in morphology (a, form II; b, form I) and Raman spectra (c) associated with the II-I transition in TA. (Note that form II and form I polymorphs in this figure were at room temperature and 95 $^{\circ}\text{C}$ respectively.)



they have no crystalline characteristics (Fig. 2a). By embedding in resin and sectioning with a microtome knife, however, the internal microstructure is revealed as an array of aligned needles (Fig. 2b and c). Making use of Ozaki and Shigeyasu's¹⁴ previous characterization of these ovoids by X-ray diffraction, and the new information on the morphology and twinning of recrystallized form I TA discussed above, we deduce that these are highly ordered arrays of twinned microcrystals with the molecular chains lying almost normal to their needle *c*-axes.

To ascertain which form is more stable at room temperature and pressure we carried out hot-stage microscopy in which single crystals and particles of TA were heated to 200 $^{\circ}\text{C}$ and any changes in morphology, opacity, or Raman spectrum noted. For form I crystals and commercial particles no changes were observed. Form II crystals (produced in-house as described earlier), however, underwent a spontaneous morphological change from rhombic to rectangular at room temperatures between 80 and 100 $^{\circ}\text{C}$. This was accompanied by a change in Raman spectrum from that of form II to that of form I. These effects are shown in Fig. 3. From these data and SR Laue single-crystal X-ray diffraction patterns of a form II crystal recorded at 30 and 150 $^{\circ}\text{C}$, we are able to state categorically that form II is the more stable form at room temperature and pressure. The morphological change corresponds to a transformation to the form I structure in which the crystal *c*-axis and chain directions are preserved but which takes place via chain sliding and rotation. This is shown in Fig. 4. We noted that for many crystals this transformation was accompanied by the release of sufficient mechanical energy to cause the crystal to jump off the microscope stage. In these untwinned crystals the phase change was found to be



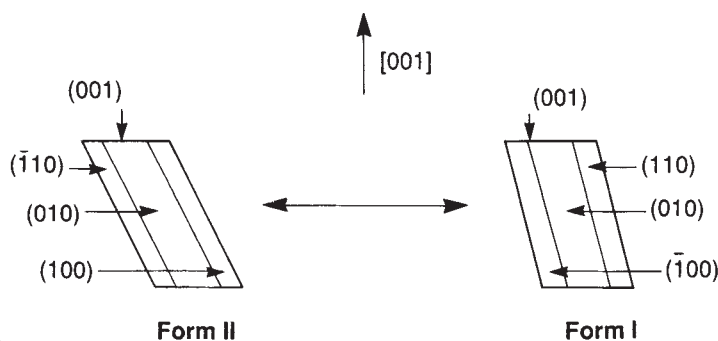
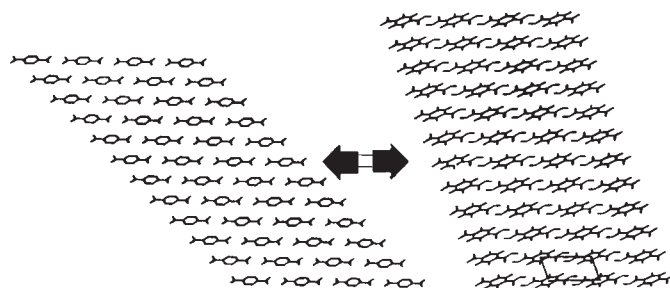


FIG. 4 Simultaneous changes in crystal morphology and structure occurring during the form II-form I transition in TA. (Hydrogen positions were only determined for the form I structure.)



reversible on cooling to $\sim 30^\circ\text{C}$ although many crystals retained the form I structure, only reverting to form II upon application of slight stress using a metal point. The phase transition appears to be martensitic in nature: no phase boundaries are present in transforming crystals and the extent of transformation appears to be a function only of temperature.

These data lead us to the surprising result that all commercial TA ovoids exist in the form I structure which is metastable at room temperature. Given that these materials have indefinite physical stability, how are we to rationalize this observation? As early as the end of the last century Ostwald¹⁵ understood that in many cases metastable polymorphs, although unfavoured thermodynamically, could appear for kinetic reasons. When such phases are isolated they will ultimately undergo a solid-state transformation to a stable phase, a process involving nucleation and growth of the new structure within the metastable crystal. The kinetics of such processes have been described^{16,18} and are well exemplified by the case of ammonium nitrate¹⁹. Accordingly, the persistence of thermodynamically metastable structures has been attributed to the low rates of kinetic processes, particularly nucleation, in the solid state. For the observations reported here such kinetic stabilization is clearly not an appropriate explanation. But taking all the structural data together the following, previously unobserved, mechanism may be proposed. When a form I crystal reverts to the form II structure a morpho-

logical change from orthogonal to rhombic occurs as the molecular chains slide and rotate into their new positions (Fig. 4). In ordered arrays of crystals twinned normal to the molecular chains (Fig. 2), as the commercial products are, both the chain movement and the associated morphological change are prevented in any one crystal by the presence of adjacent twins.

Thus, rather like a deck of cards, the stabilization of these form I particles is conferred on them collectively; phase transition would require a concerted inter-crystal process the probability of which is so low that form I is stabilized indefinitely. A test of this hypothesis would be to reduce particles to their constituent untwinned microcrystals which should then be free to undergo spontaneous transformation. Accordingly we subjected commercial TA to ultrasound in aqueous suspension. As expected, the Raman spectra indicated a marked increase in the proportion of form II present. It is our belief that this is the first time that a structurally rather than a kinetically based stabilization mechanism has been observed. Recent studies, in which both twinning²⁰ and mosaic ordering²¹ of crystals have resulted from the addition of stereo-specific additives to the crystallizing medium, lead to the view that this principle is capable of extension to other polymorphic systems; of particular interest are those systems in which the required physical properties reside in a metastable structure. □

Received 2 July; accepted 11 October 1993.

- Verma, A. R. & Krishna, P. *Polymorphism and Polytypism in Crystals* 33–60 (Wiley, New York, 1966).
- Bernstein, J. & Hagler, A. T. *J. Am. chem. Soc.* **100**, 673–681 (1978).
- Bischoff, J. L. & Fyfe, W. S. *Am. J. Sci.* **266**, 65–79 (1968).
- Keller, A., Goldbeck-Wood, G. & Hikosaka, M. *J. chem. Soc., Faraday Disc.*, No. 95 (in the press).
- Hattori, K. & McCrone, W. *Analyt. Chem.* **28**, 1791–1793 (1956).
- Hall, S. R. et al. *J. Cryst. Growth* **79**, 745–751 (1986).
- Subotic, B. et al. *Zeolites* **2**, 135–142 (1982).
- Davey, R. J. in *Crystal Growth in Science and Technology* NATO ASI Ser. (Eds Arend, H. & Hulliger, J.) 217–224 (Plenum, New York, 1989).
- Bailey, M. & Brown, C. J. *Acta crystallogr.* **22**, 387–391 (1967).
- Berkovitch-Yellin, Z. & Leiserowitz, L. *J. Am. chem. Soc.* **104**, 4052–4064 (1982).

- Gerasimov, V. P., Zharikov, N. K., Korobkov, V. S., Ovchinnikov, I. V. & Rud, L. V. *Zh. Prikl. Spektroskopii* **34**, 308–311 (1981).
- Saska, M. & Myerson, A. S. *Cryst. Res. Technol.* **20**, 201–208 (1985).
- Bhat, H. L., Clarke, S. M., El Korashy, A. & Roberts, K. J. *J. appl. Crystallogr.* **23**, 545–549 (1990).
- Ozaki, T. & Shigeyasu, M. *Maruzen Sekiyu Gihō* **20**, 81–90 (1975).
- Ostwald, W. Z. *phys. Chem.* **22**, 289–330 (1897).
- Avrami, M. *J. chem. Phys.* **7**, 1103–1112 (1939).
- Cardew, P. T., Davey, R. J. & Ruddick, A. J. *J. chem. Soc. Faraday Trans.* **80**, 659–668 (1984).
- Mnjukh, Ju. V. *Molec. Cryst. Liq. Cryst.* **52**, 201–218 (1979).
- Davey, R. J. et al. *J. Phys. D* **24**, 176–185 (1991).
- Davey, R. J. et al. *J. chem. Soc., Faraday Trans.* **88**, 3461–3466 (1992).
- Davey, R. J., Harding, M. M. & Rule, R. J. *J. Cryst. Growth* **144**, 7–12 (1991).

Insensitivity of global warming potentials to carbon dioxide emission scenarios

Ken Caldeira* & James F. Kastling

Earth System Science Center, 248 Deike Building,
Pennsylvania State University, University Park,
Pennsylvania 16802, USA

GLOBAL warming potentials for radiatively active trace gases (such as methane and chlorofluorocarbons) have generally been expressed^{1,2} relative to the time-integrated climate forcing per unit emission of carbon dioxide. Previous attempts to estimate the integrated climate forcing per unit CO₂ emitted have focused on perturbations to steady-state conditions in carbon-cycle models. But for non-steady-state conditions, the integrated climate forcing from a CO₂ perturbation depends both on the initial conditions and on future atmospheric CO₂ concentrations. As atmospheric CO₂ concentrations increase, the radiative forcing per unit CO₂ emitted will become smaller because the strongest absorption bands will already be saturated. At the same time, higher concentrations of dissolved carbon in the surface ocean will reduce the ocean's ability to absorb excess CO₂ from the atmosphere. Each of these effects taken alone would affect the climate forcing from a pulse of emitted CO₂ by a factor of three or more; but here we show that, taken together, they compensate for each other. The net result is that the global warming potential of CO₂ relative to other radiatively active trace gases is nearly independent of the CO₂ emission scenario. Thus, the concept of the global warming potential remains useful, despite the nonlinearities in the climate system and uncertainties in future emissions.

Measures of global warming potential (GWP) have typically been expressed^{1,2} as the time-integrated commitment to climate forcing from the instantaneous release of 1 kg of some trace gas *i* (ICF_{*i*}), divided by the corresponding value for CO₂

$$\text{GWP} = \text{ICF}_i / \text{ICF}_{\text{CO}_2} \quad (1)$$

The integrated climate forcing for trace gas *i* (ICF_{*i*}) may be defined as

$$\text{ICF}_i = \int_0^n a_i c_i dt \quad (2)$$

where *a_i* (W m⁻² p.p.m.⁻¹) is the instantaneous radiative forcing due to a unit increase in the atmospheric concentration (p.p.m.) of trace gas *i*, *c_i* (p.p.m.) is the concentration of trace gas *i* remaining in the atmosphere at time *t* (yr) after its release and *n* is the number of years over which the calculation is performed. The Intergovernmental Panel on Climate Change¹ neglected the dependence of both the radiative (*a_i*) and the concentration (*c_i*) terms on background trace-gas concentration, although they recognized their possible importance. In this work, we calculate these terms for CO₂, for several future atmospheric CO₂ scenarios.

The concentration CO₂ remaining in the atmosphere (*c*_{CO₂}, in p.p.m.) at time *t* following a CO₂ release may be defined as the amount of CO₂ in the atmosphere at time *t* minus the amount of CO₂ that would have been in the atmosphere at time *t* in the absence of that release. Although *c*_{CO₂} should, in principle, depend on the rates of CO₂ exchange with both the terrestrial biosphere and the ocean, we follow previous investigators^{1,2} in considering only the ocean. (The terrestrial biosphere can act as either a sink or source of CO₂, depending on future land-use

TABLE 1 Coefficients for fit to impulse-response functions

Scenario	<i>k</i> ₀	<i>k</i> ₁	<i>k</i> ₂	<i>k</i> ₃	<i>k</i> ₄	<i>τ</i> ₁	<i>τ</i> ₂	<i>τ</i> ₃	<i>τ</i> ₄
1	0.265	0.546	-0.804	0.903	0.090	1789.3	45.7	25.0	1.4
2	0.280	0.498	-0.667	0.751	0.138	1814.2	74.8	36.0	2.6
3	0.345	0.236	0.306	0.060	0.054	738.1	20.5	2.6	1.4
4	0.297	0.321	0.266	0.083	0.033	335.8	18.4	2.8	0.8
5	0.187	0.261	0.286	0.141	0.125	316.5	88.3	12.2	2.6
6	0.152	0.289	0.275	0.202	0.082	242.4	46.1	9.4	1.4

These coefficients are as described in equation (3) and represent an exponential least-squares fit to the impulse-response function shown in Fig. 2a and derived from the box-diffusion model described in the text.

practices and the uncertain effects of CO₂ fertilization on plant growth in natural ecosystems.) In general, *c*_{CO₂} depends on future atmospheric CO₂ concentrations, because the partial pressure of CO₂ in the ocean surface is a nonlinear function of surface total dissolved inorganic carbon concentration ([Σ CO₂]). In this work, we use a box-diffusion ocean model³ similar to that of Siegenthaler⁴, using the carbonate and borate chemistry described by Stumm and Morgan⁵, the gas-exchange coefficient used by Maier-Reimer and Hasselmann⁶, and an eddy diffusion constant derived from bomb ¹⁴C tuning⁴. Model results³ for the fraction of a CO₂ pulse remaining in the atmosphere for pulses of 0.25, 1 and 3 PAL (where 1 PAL represents a pre-industrial atmospheric level of 265 p.p.m.) injected into a steady-state atmosphere are similar to the results of Maier-Reimer and Hasselmann⁶; the r.m.s. and maximum differences are <0.01 <0.04 respectively. This model is appropriate only on timescales <1,000 yr, when interaction with sediments and rock cycles is of secondary importance.

To see how *c*_{CO₂} depends on the CO₂ emission scenario, we investigated six scenarios defined in terms of atmospheric CO₂ content as a function of time (Fig. 1a). Scenarios 1 and 2 represent the CO₂ concentrations calculated by the model³ for two logistic fossil-fuel burning scenarios (refs 6 and 7), both with total recoverable reserves of 5,000 GtC, and 1990 CO₂ release set to 6 GtC. In these cases the initial growth in fossil-fuel burning are 4.3 and 2.35% per year, respectively. Scenarios 3, 4 and 5 represent stabilization scenarios in which atmospheric CO₂ content is stabilized at 750, 550 and 350 p.p.m. by years 2250, 2150 and 2150, respectively. In these three cases, the 1990 growth rate in atmospheric CO₂ content is 1.66 p.p.m. yr⁻¹ (equal to the growth rate between 1980 and 1990), and the growth rate at the stabilization date is zero. From these conditions, a unique third-order polynomial was determined and used to specify the transition to the stabilized atmospheric CO₂ content. For scenarios 1–5, the initial 1990 state is determined by initializing the box-diffusion model at year 1765, in a steady state with an atmospheric *p*CO₂ of 279 p.p.m., and then integrating forward to 1990 according to the historical CO₂ data given by Shine *et al.*¹. Scenario 6 represents a steady-state 279 p.p.m. atmosphere. The CO₂-emission emissions (as a function of time) required to produce the six scenarios are shown in Fig. 1b.

For each of these scenarios, an additional small pulse of CO₂ was injected into the atmosphere in year 1990. The impulse-response function (*c*_{CO₂}) for each of these scenarios is presented in Fig. 2a. The coefficients for least-square curve fits to these results, in the form

$$c_{\text{CO}_2} = k_0 + \sum_{j=1}^4 k_j \exp(-t/\tau_j) \quad (3)$$

are shown in Table 1.

The radiative forcing from CO₂ (Δ*F*_{CO₂}) is a function of CO₂ concentration (*C*)

$$\Delta F_{\text{CO}_2} = 6.3 \ln(C/C_0) \quad (4)$$

* Present address: Global Climate Research Division, Lawrence Livermore National Laboratory, PO Box 808, L-256, Livermore, California 94550, USA.

TABLE 2 Model results

Scenario	Description	Fraction of CO ₂ remaining in the atmosphere			Fraction of initial climate forcing			Time-integrated climate forcing (relative to scenario 6)		
		20	100	500	20	100	500	20	100	500
1	Logistic 4.3% yr ⁻¹	0.69	0.71	0.68	0.60	0.17	0.17	0.87	0.72	0.63
2	Logistic 2.35% yr ⁻¹	0.69	0.62	0.66	0.62	0.26	0.16	0.88	0.82	0.68
3	Stabilize at 750 p.p.m.	0.69	0.55	0.46	0.63	0.35	0.22	1.02	1.02	1.04
4	Stabilize at 550 p.p.m.	0.69	0.54	0.37	0.63	0.37	0.24	1.03	1.03	1.14
5	Stabilize at 350 p.p.m.	0.69	0.47	0.24	0.64	0.45	0.24	1.03	1.11	1.26
6	Steady-state at 279 p.p.m.	0.62	0.37	0.19	0.62	0.37	0.19	1	1	1

The fraction of CO₂ remaining in the atmosphere is c_{CO_2} . The fraction of the initial climate forcing is $a_{CO_2} c_{CO_2}$ divided by the value immediately after the CO₂ emission. Scenario 6 values for the time-integrated climate forcing (ICF_{CO₂}) are 8.5×10^6 , 3.0×10^7 and 8.6×10^7 J m⁻² p.p.m.⁻¹ (1 p.p.m. = 2.1 GtC), for 20, 100 and 500 yr time horizons, respectively, calculated using the radiative forcing at year 1990.

where C_0 is in units of p.p.m. CO₂, and ΔF_{CO_2} is in W m⁻² (refs 1, 8). The logarithmic form of this relation arises because the strongest absorption bands of CO₂ (the 15- μ m band in particular) are already saturated near band centre; so, the increased absorption resulting from the addition of CO₂ to the atmosphere must take place in the wings of these bands. The change in radiative forcing per unit CO₂ (a_{CO_2}), due to a small change in atmospheric CO₂ concentration is

$$a_{CO_2} = \frac{d\Delta F_{CO_2}}{dC} = \frac{6.3}{C} \quad (5)$$

in units of W m⁻² p.p.m.⁻¹, with C in units of p.p.m.

Substituting equations (3) and (5) into equation (2) allows us to calculate the integrated climate forcing from the release of

1 kg CO₂ for each of the scenarios described above (Table 2). The logistic fossil-fuel scenarios extend the value of ΔF_{CO_2} outside the range of its parameterization (<1,000 p.p.m., refs 1, 8). Nevertheless, radiative-convective model results⁹ indicate that the logarithmic form of this equation is appropriate in the range used here.

The results of the integrated climate forcing calculation (Table 2) show that the climate forcing from a CO₂ emission is independent of the CO₂-emissions scenario to $\pm 33\%$. This is despite a factor of three difference in both the radiative forcing (a_{CO_2}) and the airborne fraction (c_{CO_2}) of the emitted CO₂. Equation (5) shows that the change in climate forcing from an increase in atmospheric CO₂ is inversely proportional to the atmospheric CO₂ concentration. But as the ocean and atmosphere approach

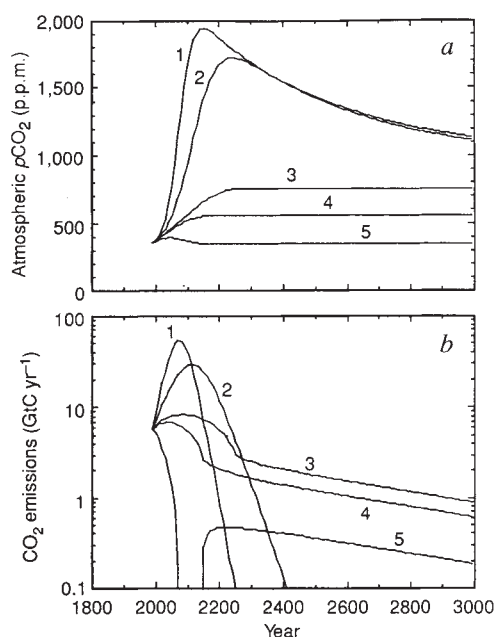


FIG. 1 a, Atmospheric CO₂ scenarios for which non-steady-state impulse-response functions were calculated. b, Total fossil fuel plus net biosphere CO₂ emissions calculated by the box-diffusion model, corresponding to each of these atmospheric CO₂ scenarios. The logistic growth equation used to describe scenarios 1 and 2 is $dQ/dt = \mu Q(1 - Q/R)$ where dQ/dt is the CO₂ emission rate, R is the total recoverable fossil fuel (5,000 GtC), $\mu = 4.3$ and 2.35% for scenarios 1 and 2, respectively. Q was adjusted such that emissions were 6 GtC yr⁻¹ in 1990. For scenario 5, net emissions must be <0 near year 2100 to stabilize atmospheric pCO₂ at 350 p.p.m. This could occur if fossil fuel burning were greatly reduced and the biosphere acted as a net sink for CO₂.

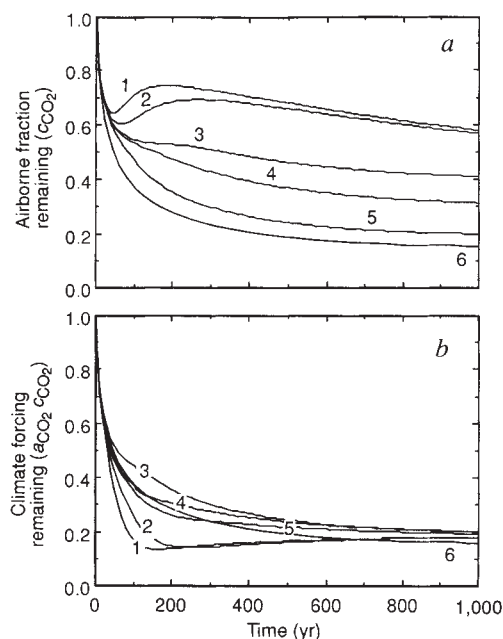


FIG. 2 a, Non-steady-state impulse-response functions corresponding to each of the atmospheric CO₂ scenarios shown in Fig. 1a. The intermediate maxima in scenarios 1 and 2 occur because present-day CO₂ emission makes the ocean less able to absorb CO₂ emitted in the future. b, Climate forcing relative to the forcing at the time of the initial atmospheric CO₂ perturbation, corresponding to each of the atmospheric CO₂ scenarios. Note that, among these scenarios, the climate forcings from a pulse of CO₂ are much more similar to each other than are the impulse-response functions. At higher CO₂ concentrations, the climate forcing is less sensitive to changes in atmospheric CO₂ and the oceans less readily absorb atmospheric CO₂. These two nonlinearities (one in radiation physics, one in marine chemistry) tend to cancel each other.

steady state (disregarding sediment interactions), the asymptotic airborne fraction for a small pulse of CO_2 emitted to the atmosphere is roughly proportional to the atmospheric CO_2 concentration³. (This is because, with higher CO_2 content, the ocean is less able to absorb atmospheric CO_2 .) Hence, for a small pulse, the airborne fraction times the additional climate forcing is nearly independent of the ambient atmospheric CO_2 concentration. This can be seen in Fig. 2b. On timescales approaching 1,000 yr, dissolution of carbonate sediments should contribute to the absorption of CO_2 by the ocean, lessening the remaining airborne fraction of the emitted CO_2 .

The results shown here imply that the concept of global warming potential may be useful despite nonlinearities in the climate system and uncertainty regarding future CO_2 emissions. It is important, however, that the high GWPs of the various anthropogenic trace gases ($>7,000$ for some CFCs) are not taken to mean that these gases are a more serious problem than CO_2 . Figure 2 and Table 2 show that $\sim 20\%$ of the climate forcing from CO_2 emitted into the atmosphere today should still be there

500 to 1,000 yr from now. This makes CO_2 much more long-lived than other anthropogenic gases, most of which have lifetimes ranging from 10 yr (for methane) to 150 yr (for nitrous oxide). Thus, climate forcing by CO_2 is more difficult to reverse on human timescales than is forcing by other greenhouse gases. Continued high rates of CO_2 emission will commit us to many centuries of warmer climate. □

Received 13 July; accepted 7 October 1993.

1. Shine, K. P., Derwent, R. G., Wuebbles, D. J. & Morcrette, J.-J. in *Climate Change—The IPCC Scientific Assessment* (eds Houghton, J. T., Jenkins, G. J. & Elphraums, J. J.) 41–68 (Cambridge Univ. Press, New York, 1990).
2. Lashof, D. A. & Ahuja, D. R. *Nature* **344**, 529–531 (1990).
3. Caldeira, K. *Global Biogeochem. Cycles* (submitted).
4. Siegenthaler, U. *J. geophys. Res.* **88**, 3599–3608 (1983).
5. Stumm, W. & Morgan, J. J. *Aquatic Chemistry* 2nd edn (Wiley, New York, 1981).
6. Maier-Reimer, E. & Hasselmann, K. *Clim. Dyn.* **2**, 63–90 (1987).
7. Killough, G. G. & Emanuel, W. R. *Tellus* **33**, 274–290 (1981).
8. Hansen, J. et al. *J. geophys. Res.* **93**, 9341–9364 (1988).
9. Kiehl, J. T. & Dickinson, R. E. *J. geophys. Res.* **92**, 2991–2998 (1987).

ACKNOWLEDGEMENTS. K.C. was supported by the US NSF.

Leaching and reconstruction at the surfaces of dissolving chain-silicate minerals

William H. Casey*, Henry R. Westrich†, Jillian F. Banfield‡, Giulio Ferruzzi* & George W. Arnold§

* Department of Land, Air and Water Resources and the Department of Geology, University of California, Davis, California 95616, USA

† Geochemistry Research and § Ion-Solids Interaction Division, Sandia National Laboratories, Albuquerque, New Mexico 87185, USA

‡ Department of Geology, University of Wisconsin, Madison, Wisconsin 53706, USA

THE pathways by which silicate minerals transform to solutes, clays and amorphous solids are relevant to a wide range of natural, industrial and even medical concerns. For example, weathered layers on silicate may have a high sorptive capacity, affecting nutrient and contamination retention in soils; less obviously, such layers on inhaled silicate grains might affect their interaction with lung tissue. Here we report the observation, in dissolution experiments on a range of chain-silicate minerals, of the formation of a near-surface amorphous region enriched in silicon and hydrogen, and depleted in other metals. Raman spectroscopy and ion-beam elemental analysis show that portions of the polymeric silicate anion in this region spontaneously reconstruct to form a network that contains four-member silicate rings and areas of incipient crystallization. If hydrolysable metals interact with the silicate anion during this reconstruction, clays and amorphous products may form directly. This process complements traditional dissolution–precipitation pathways of mineral diagenesis¹, as the silicon does not have to be present in solution before being incorporated into a growing secondary phase.

The polymeric silicate anion at the surface of most rock- and soil-minerals must protonate and dissociate to release silicic acid. Little is known about this depolymerization and there are some puzzling results. Exchange of protons for the alkaline-earth cations at the surface of chain-silicate minerals is rapid in acidic solution^{2–10}. However, the experimental fluxes of dissolved silicon from chain-silicate minerals vary considerably with mineral composition (Table 1). Wollastonite rates are much higher than other chain-silicate minerals, indicating that silicate chains are apparently unequally reactive even after divalent metals are

replaced with protons. Paradoxically, the specific area of a mineral powder sometimes increases measurably during a dissolution experiment while the flux of dissolved silica is constant or even decreases with time^{6,7}. If the mineral surface were uniformly reactive, the flux of dissolved silica would increase with the exposed surface area.

Here we test the hypothesis^{3–5} that portions of the exposed silicate anion reconstruct during dissolution when divalent met-

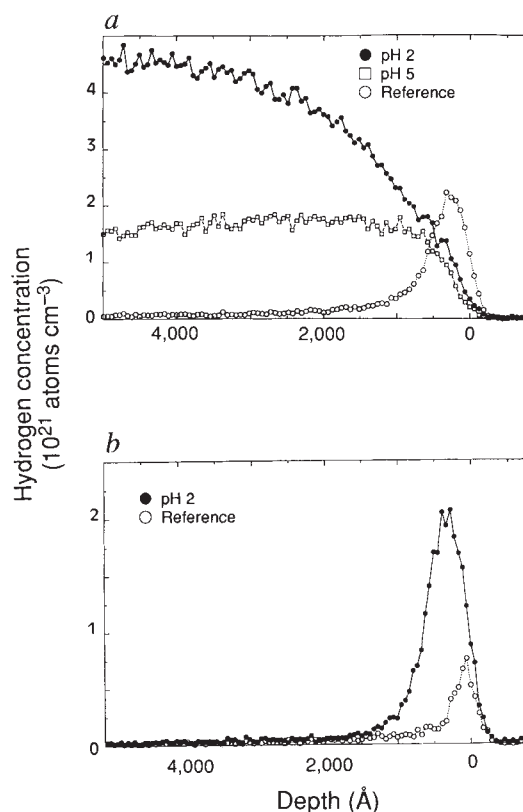


FIG. 1 Depth profiles of hydrogen concentration in a, wollastonite, and b, diopside. The analytical conditions were as reported previously^{3,4}. Crystals were suspended in large carboys of solution at fixed pH (2 and 5), ionic strength (0.1 M) and temperature (50 °C) for 522 h. Dissolved silicon concentrations were maintained below 1.5×10^{-4} M, which is well below saturation with amorphous silica.

TABLE 1 Variation of flux of dissolved silicon at pH 2 from chain-silicate minerals with mineral composition

Composition	Common name	Rate (mol cm ⁻² s ⁻¹)	Silicate structure	Reference
α -CaSiO ₃ *	Pseudowollastonite	10 ^{-8.1}	Trisilicate rings	G.F., in prepn
CaSiO ₃ †	Wollastonite	10 ^{-10.4}	Pyroxenoid chains	6
CaSiO ₃ †	Wollastonite	10 ^{-10.15}	Pyroxenoid chains	7
MgSiO ₃ †	Enstatite	10 ^{-13.9}	Pyroxene chains	10
MnSiO ₃ †	Rhodonite/pyroxmangite	10 ^{-13.8}	Pyroxenoid chains	G.F., in prepn
MnSiO ₃ *	Rhodonite	10 ^{-13.8}	Pyroxenoid chains	G.F., in prepn
CaMg(SiO ₃) ₂ †	Diopside	10 ^{-12.8}	Pyroxene chains	10
CaMg(SiO ₃) ₂ †	Diopside	10 ^{-13.95}	Pyroxene chains	26
SiO ₂ †	Amorphous silica	10 ⁻¹⁶	Tectosilicate	27
SiO ₂ †	Quartz	10 ^{-17.2} –10 ^{-16.2}	Tectosilicate	28

Dissolved silicon fluxes are from * synthetic and † natural metasilicate minerals and silica polymorphs in highly undersaturated chloride solutions at 23–25 °C. Results from refs 6, 7 and 10 were extrapolated from experiments in the pH range of 3–8. The natural MnSiO₃ is an intimate intergrowth of two phases that are polymorphs; they differ in the repeat interval along the pyroxenoid chains.

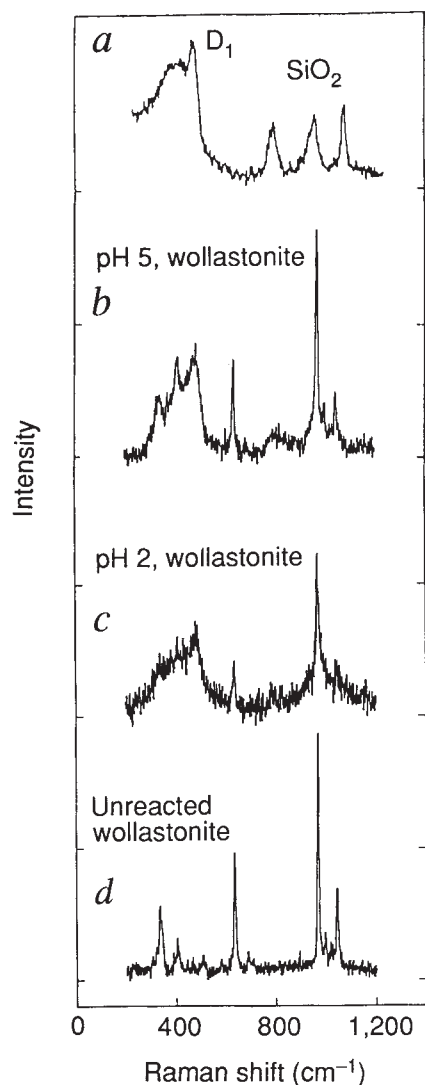


FIG. 2 Raman spectra from wollastonite (b, c, d) compared with a published spectrum for amorphous silica produced by the sol-gel method (a)¹³. Wollastonite was reacted as for Fig. 1. The analytical spot size allows for 1 μ m horizontal resolution and a sampling depth of 5–10 μ m. Raman spectra were collected in flowing helium using the 457.9 nm excitation line of an Ar⁺ laser.

als are selectively exchanged for protons. We chose to test this hypothesis with dissolution experiments on iron-free chain-silicate minerals because the structures have no covalent cross-links to metals other than silicon, the metal valence states are stable, and divalent metals are selectively leached from the near-surface regions over a wide pH range^{5–11}. We combine Raman spectroscopy with ion-beam analyses of single crystals and dissolution experiments on mineral powders to infer changes in the protonation and configuration of the silicate anion in the leached portion of the mineral surface.

Elemental depth profiles (Fig. 1) indicate that ion exchange and hydration proceed to considerable depth below the mineral-solution interface, as shown previously². The near-surface region of wollastonite is enriched in hydrogen to depths greater than several thousands of ångströms, whereas diopside is affected to depths of $\sim 1,000$ Å for a similar period (Fig. 1). These near-surface regions are enriched in silicon and depleted of calcium and magnesium relative to unreacted materials, as determined by Rutherford backscattering spectra (not shown) and analytical transmission electron microscopy (not shown).

Surprisingly, hydrogen concentration in wollastonite reacted at pH 5 is nearly constant with depth but differs considerably from that reacted at pH 2 (Fig. 1). The large difference in concentration leads us to propose that non-bridging oxygens are eliminated to form siloxane bridges



after exchange of protons for calcium ions. The water produced is presumably lost during drying and was thus not analysed.

Raman spectra of the reacted wollastonite surface (Fig. 2) confirm our hypothesis. The most significant difference between spectra for reacted and unreacted samples is a broad band in the 350–500 cm⁻¹ region that is usually interpreted as bending/stretching of silicate tetrahedra in a tectosilicate fabric^{12–15}. The distinct peak near 490 cm⁻¹ (*D*₁) arises from the breathing or stretching modes of vibrationally decoupled fourfold silicate rings¹⁵. The spectra from reacted surfaces also have features in common with unreacted wollastonite. In particular, the 637 and 950–970 cm⁻¹ Raman lines are characteristic of bending/stretching modes of silicate chains^{12,16} that may exist in the reacted region or as unreacted wollastonite which is being sampled through the reacted near-surface region.

Wollastonite is exceptional in that the rates of reaction are much more rapid than for other chain-silicate minerals and the incongruence is profound. High-resolution transmission electron micrographs show this reacted surface on other chain-silicate phases, even in cases where it is only tens to hundreds of ångströms thick, and provide evidence for the reaction shown in equation (1) at almost neutral pH conditions. The near-surface region on MnSiO₃ (Fig. 3) after reaction at pH 7.1 for 2,225 h,

for example, is amorphous to electron diffraction and resembles amorphous silica (see also ref. 5). Manganese is eliminated from the near-surface region by exchange for protons, and small spheres of crystallizing silica are indicated by lattice fringes. The dissolved silicon concentrations in the experiment that produced this surface were always less than 4×10^{-5} M, which is below saturation with all silica polymorphs.

If the restructured portion of the mineral surface were as unreactive as vitreous silica (Table 1), considerable surface area could be added without enhancing the flux of dissolved silica. The extent of restructuring certainly varies with solution composition (particularly pH), mineral structure and composition¹¹, as indicated by the dramatic differences in leaching depths for different minerals in similar conditions (Fig. 1). To demonstrate further

the decreased reactivity, we rinsed 0.1 g wollastonite powder for 10 min with a flowing 0.01 M HCl solution. The BET-measured surface area increases from $0.4 \text{ m}^2 \text{ g}^{-1}$ to $1.6 \text{ m}^2 \text{ g}^{-1}$, whereas the measured flux of silicic acid decreases by over an order of magnitude (G.F. *et al.*, manuscript in preparation). Clearly a portion of the exposed silicate anion is restructuring to a less reactive form with time.

Both the rate and congruency of dissolution relate closely to the silicate anion structure and mineral composition. The silicate anion in pseudowollastonite, for example, exists as three-membered rings and not as chains. Pseudowollastonite dissolves completely congruently, whereas wollastonite dissolves more slowly (Table 1) and incongruently. Within the chain-silicate class of minerals, the strengths of bonds from silicate chains to divalent metals in bulk structure are important. Magnesium bonds more strongly to silicate chains than calcium¹⁰. Correspondingly, diopside ($\text{CaMgSi}_2\text{O}_6$) dissolves more slowly, and with much less extensive leaching, than wollastonite (CaSiO_3); we have not yet found any convincing evidence for extensive leaching of enstatite (MgSiO_3). Subtle differences in chain configuration, however, are relatively unimportant. The dissolution rates of rhodonite and pyroxmangite (both MnSiO_3) are virtually equal (Table 1), but the structures differ in the repeat interval along the pyroxenoid chains.

The observation that the silicate anion restructures during mineral leaching links results of dissolution experiments to studies of natural weathering. Primary and secondary minerals are sometimes so intergrown in nature that it is suggested^{17, 20} that polymeric fragments of the primary phase are incorporated directly into a growing secondary mineral. The abundance of water in a dissolution experiment is thought^{17, 18} to preclude such topotactic conversion leading rather to complete depolymerization of the exposed silicate polymer and release of dissolved monomers. Silicate glasses^{14, 21} and calc-silicate minerals in cement²² also react *in situ* to form an assemblage of secondary silicates and amorphous materials. These transformations are caused by hydration of the metals followed by hydrolysis and rearrangement of the silicate fabric to eliminate strain.

We have shown that silicon need have no residence time in solution as silicic acid before it is incorporated into a solid reaction product at the surface of a mineral. One can imagine that, if this restructuring is affected by interaction with adsorbed metals, it leads to crystallization of secondary minerals by what is essentially a gel-mediated path. Smectite clays, for example, are easily synthesized by crystallizing silicate gels²³. Many minerals dissolve incongruently at pH conditions that are nearly neutral^{24, 25} and the resulting leached surfaces are ideal for such a reaction. □

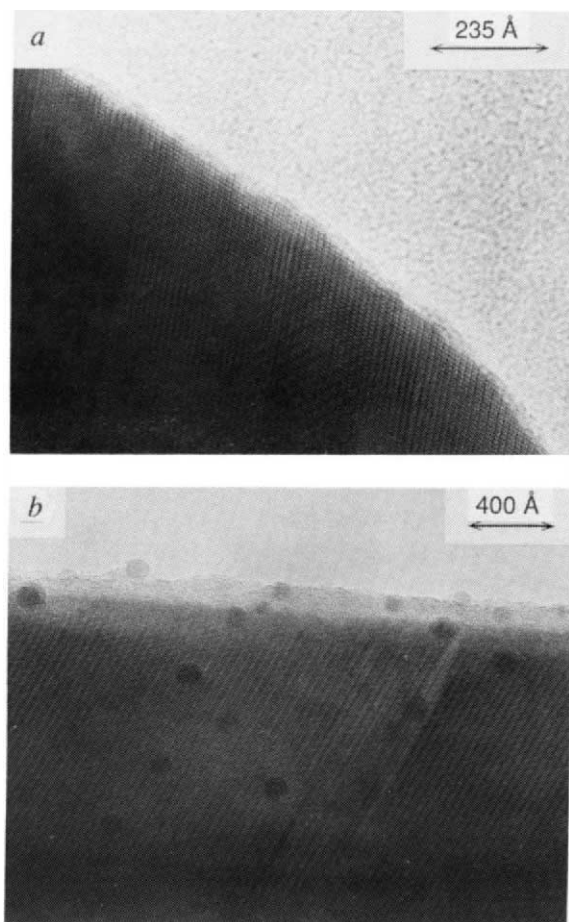


FIG. 3 High-resolution-transmission electron micrograph of an unreacted (a) and reacted (b) grains of MnSiO_3 . Note that lattice fringes in MnSiO_3 extend to within 10–20 Å of the surface of the unreacted sample (a) but are missing from the near-surface region of the reacted sample (b). The leached layer in (b) is amorphous to electron diffraction to depths of 2–400 Å. This region corresponds to the silicon- and hydrogen-enriched layer that is detectable by ion-beam analysis (as in Fig. 1 for example). The dark spheres in b are areas of condensation and incipient crystallization of the leached surface; because the entire particle is reacted, spheres appear to form in bulk mineral as well as at the particle edge. To prepare the sample, grains of powder were dispersed directly onto a holey carbon grid for imaging. Transmission electron micrographs were recorded on a Philips 400T transmission electron-microscope operated at 120 keV. After this dissolution experiment, sample grains were rinsed with water, ethanol and acetone, followed by drying at ambient temperature.

Received 26 July; accepted 28 September 1993.

1. Kittrick, J. A. (ed.) *Soil Mineral Weathering* (Van Nostrand Reinhold, New York, 1986).
2. Petit, J.-C., Della Mea, G., Dran, J.-C., Schott, J. & Berner, R. A. *Nature* **325**, 705–706 (1987).
3. Casey, W. H., Westrich, H. R. & Arnold, G. W. *Geochim. cosmochim. Acta* **52**, 2795–2807 (1988).
4. Casey, W. H., Westrich, H. R., Arnold, G. W. & Banfield, J. F. *Geochim. cosmochim. Acta* **53**, 821–832 (1989).
5. Murphy, W. M. & Helgeson, H. C. *Am. J. Sci.* **289**, 3137–3153 (1989).
6. Bailey, A. & Reesman, A. L. *Am. J. Sci.* **271**, 464–472 (1971).
7. Rimstidt, J. D. & Dove, P. M. *Geochim. cosmochim. Acta* **50**, 2509–2516 (1986).
8. Luce, R. W., Bartlett, R. W. & Parks, G. A. *Geochim. cosmochim. Acta* **36**, 35–50 (1972).
9. Murphy, W. M. & Helgeson, H. C. *Geochim. cosmochim. Acta* **51**, 3137–3153 (1987).
10. Schott, J., Berner, R. A. & Sjöberg, E. L. *Geochim. cosmochim. Acta* **45**, 2123–2135 (1981).
11. Casey, W. H. & Bunker, B. in *Mineral–Water Interface Geochemistry* (eds Hochella, M. F. Jr & White, A. F.) 397–426 (Miner. Soc. Am., Washington DC, 1990).
12. Furukawa, T., Fox, K. E. & White, W. B. *J. Chem. Phys.* **75**, 3226–3237 (1981).
13. Brinker, C. J., Tallant, D. R., Roth, E. P. & Ashley, C. S. *J. non-cryst. Solids* **82**, 117 (1986).
14. Bunker, B. C., Tallant, D. R., Headley, T. J., Turner, G. L. & Kirkpatrick, R. J. *J. Phys. Chem. Glasses* **29**, 106–120 (1988).
15. Galeener, F. L. *J. non-cryst. Sol.* **53**, 2823–2830 (1989) (1982).
16. Conjeaud, M. & Boyer, H. *Cem. Concr. Res.* **10**, 61–70 (1980).
17. Casey, W. H., Banfield, J. F., Westrich, H. R. & McLaughlin, L. *Chem. Geol.* **105**, 1–15 (1993).
18. Eggleton, R. A. & Boland, J. N. *Clays Clay Miner.* **30**, 11–20 (1982).
19. Banfield, J. F., Jones, B. F. & Veblen, D. R. *Geochim. cosmochim. Acta* **55**, 2781–2793 (1991).
20. Banfield, J. F. & Eggleton, R. A. *Clays Clay Miner.* **38**, 77–89 (1990).

21. Mazer, J. J., Bates, J. K., Bradley, J. P., Bradley, C. R. & Stevenson, C. M. *Nature* **357**, 573–576 (1992).
22. Taylor, H. F. W. *Cement Chemistry* 123–166 (Academic, New York, 1990).
23. Güven, N. in *Hydrous Phyllosilicates* (ed. Bailey, S. W.) 531–535 (Miner. Soc. Am., Washington DC, 1988).
24. Bales, R. C. & Morgan, J. J. *Geochim. cosmochim. Acta* **49**, 2281–2288 (1985).
25. Hume, L. A. & Rimstidt, J. D. *Am. Miner.* **77**, 1125–1128 (1992).
26. Knauss, K. G., Nguyen, S. N. & Weed, H. C. *Geochim. cosmochim. Acta* **57**, 285–294 (1993).
27. Wirth, G. S. & Gieskes, J. M. *J. Coll. Interf. Sci.* **68**, 492–500 (1979).
28. Dove, P. N. & Elston, S. F. *Geochim. cosmochim. Acta* **56**, 4147–4156 (1992).

ACKNOWLEDGEMENTS. This letter was improved as a result of reviews by W. M. Murphy and discussions with W. B. White, D. Bunker and D. Tallant. Support for this research was from the NSF, the US Nuclear Regulatory Commission and the DOE.

Three generations of diamonds from old continental mantle

S. H. Richardson*, J. W. Harris† & J. J. Gurney*

*Department of Geological Sciences, University of Cape Town, Rondebosch 7700, South Africa

†Department of Geology and Applied Geology, University of Glasgow, Glasgow G12 8QQ, UK

INCLUSION-BEARING diamonds erupted by kimberlites are time capsules from the sub-continental mantle. Diamonds with peridotitic mineral inclusions ('peridotitic diamonds') from Cretaceous kimberlites in southern Africa have resided in stable mantle beneath the Kaapvaal craton since the Archaean¹. Diamonds with eclogitic inclusions ('eclogitic diamonds') reflect episodic mantle events during the Proterozoic^{2,3}. Here we discuss and present isotope data for a further category of diamonds from the Proterozoic Premier kimberlite, in which peridotitic inclusion minerals have compositions reflecting an origin in both harzburgitic (clinopyroxene-free) and lherzolitic (clinopyroxene-bearing) assemblages. The harzburgitic garnet inclusions and their counterparts found as isolated minerals ('macrocrysts') in the kimberlite host rock⁴ have neodymium and strontium isotope signatures consistent with an Archaean age (>3,000 Myr) and metasomatized mantle source. Lherzolitic garnet and clinopyroxene inclusions yield a preferred Sm–Nd

isochron age of 1,930 Myr, which is ~100 Myr less than that of the adjacent Bushveld Complex⁵, suggesting a link analogous to that between harzburgitic diamond formation and komatiitic magmatism in the Archaean¹. The final generation of diamonds, those with eclogitic inclusions, were formed ~1,150 Myr ago, just before kimberlite emplacement⁶.

The Premier kimberlite is located in the Central Terrains of the Archaean Kaapvaal craton⁷, with a preferred emplacement age of $1,180 \pm 30$ Myr (refs 2, 6, 8–11). Underground mining operations reveal that the kimberlite, which is cut by a 1,100-Myr-old gabbroic sill¹², penetrates a norite phase of the Bushveld Complex (P. Bartlett, personal communication). The Bushveld Complex is among the largest layered intrusions known, comprising a sequence of basic rocks 7–9 km thick, with an areal extent of 66,000 km² (the Rustenburg Layered Suite) and subsequent granites¹³. Although the Premier mine is best known as the source of the world's largest gem diamonds, its most significant feature for us is the comprehensive suite of inclusions (from separate diamonds) with major element compositions covering the worldwide range for such inclusions. About 40% of the inclusions are peridotitic and 60% eclogitic¹⁴. Furthermore, the peridotitic olivines, orthopyroxenes and garnets are bimodal in composition¹⁴, reflecting a more magnesian, clinopyroxene-free harzburgitic assemblage and a more calcic, clinopyroxene-bearing lherzolitic assemblage, a distinction that is also apparent in the carbon isotope compositions¹⁵ of the host diamonds. Eclogitic diamonds from Premier yielded a Sm–Nd isochron age of $1,150 \pm 60$ Myr (ref. 6), which is within error of the kimberlite age ($1,180 \pm 30$ Myr; refs 8, 9), yet consistent with their crystallization at least 1–10 Myr before emplacement, as required by their nitrogen aggregation state⁶. Thus, this locality provides the opportunity to characterize the sources and sequence of harzburgitic, lherzolitic and eclogitic diamond formation.

Inclusions averaging 200 µm in diameter were recovered from 2-mm-sized diamonds using a steel cracker. Varieties of garnet and clinopyroxene, the only inclusion phases with enough radiogenic Nd and Sr for dating, were identifiable on the basis of both colour and major element composition. Individual inclusions (containing only 20–200 pg Nd per grain) were then grouped in categories on the basis of Ca, Cr, Ti and Na contents (Table 1) to obtain sufficient quantities of Nd and Sr for isotope analysis.

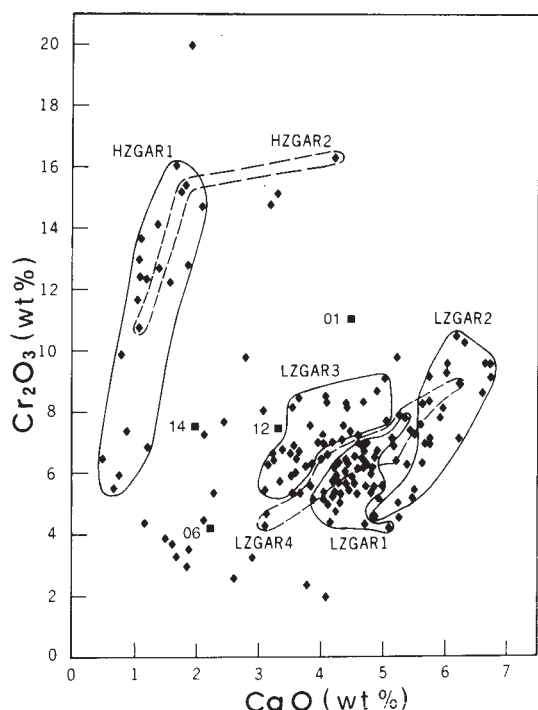


FIG. 1 Distribution of CaO and Cr₂O₃ in peridotitic garnets occurring as inclusions in diamonds (diamond symbols) and as discrete macrocryst grains in heavy mineral concentrate (square symbols; sample numbers as in Table 1) from the Premier kimberlite. Fields are drawn for the diamond inclusion composites defined in Tables 1 and 2. Unassigned (outside fields) and compromised (touching clinopyroxene or altered) inclusions were not used for isotope analysis.

TABLE 1 Major-element composition of peridotitic garnet and clinopyroxene inclusions and macrocrysts from the Premier kimberlite

	Macrocryst grains				Inclusions in diamonds							
	MGAR01	MGAR06	MGAR12	MGAR14	HZGAR1	HZGAR2	LZGAR1	LZGAR2	LZGAR3	LZGAR4	LZCPX1	LZCPX2
SiO ₂	41.21	42.85	42.59	42.56	42.09	41.16	42.18	41.75	42.37	42.12	55.75	55.47
TiO ₂	0.22	ND	0.12	0.06	ND	0.24	0.28	0.10	0.04	0.14	0.22	ND
Al ₂ O ₃	15.59	21.33	18.07	18.35	15.91	12.83	18.73	17.52	18.19	18.44	3.11	1.16
Cr ₂ O ₃	11.00	4.25	7.48	7.55	11.25	14.13	5.68	7.68	7.11	6.42	2.19	1.30
FeO	6.26	5.65	5.50	5.10	4.54	4.86	6.58	6.26	6.06	6.42	2.69	2.82
MnO	0.21	0.26	0.28	0.20	0.19	0.20	0.29	0.25	0.24	0.27	ND	ND
MgO	20.81	23.39	22.15	23.61	24.40	23.36	21.25	20.19	21.98	21.15	17.07	19.60
CaO	4.48	2.23	3.33	2.01	1.22	2.37	4.48	5.86	3.87	4.58	15.90	18.42
Na ₂ O	ND	ND	ND	ND	ND	ND	0.07	ND	ND	ND	2.91	1.01
K ₂ O	—	—	—	—	—	—	—	—	—	—	0.05	0.13
Total	99.78	99.96	99.52	99.44	99.60	99.15	99.54	99.61	99.86	99.54	99.89	99.91

Concentrations in weight% as determined by electron microprobe (ND, not detected) on splits of grains used for isotope analysis. Macrocryst (M) grains were selected on the basis of colour from kimberlite heavy-mineral concentrate (Anglo American Research Laboratories, sample M92/0534). Inclusion-bearing diamonds were selected from the -6+5 diamond sieve fraction (1.8–2.2 mm diameter) of general mine production. Inclusion data are averages for the 3- to 32-diamond composites in Table 2. Inclusion garnet (GAR) composites are identified as harzburgitic (HZ) or lherzolitic (LZ) on the basis of the respective absence or coexistence of clinopyroxene (CPX) in multiminerally-inclusion-bearing diamonds.

Bright purple harzburgitic garnets have substantially lower Ca and higher Cr contents than lherzolitic garnets (Fig. 1). Furthermore, harzburgitic garnets are generally characterized by Ti contents below electron-microprobe detection limits (category HZGAR1). Three harzburgitic garnets with significant Ti contents (0.2–0.3 TiO₂) were grouped separately (HZGAR2). The best-defined lherzolitic garnet category (LZGAR1) has significant Ti and Na contents (up to 0.9% TiO₂, 0.2% Na₂O) and a distinctive reddish tinge. In rare multiminerally-inclusion-bearing diamonds, this variety of garnet coexists with bright emerald-

green clinopyroxene (LZCPX1), which has correspondingly higher Ti, Al, Cr and Na contents than the pale emerald-green clinopyroxene (LZCPX2) that coexists with more calcic pale-purple lherzolitic garnets (LZGAR2). A further two lherzolitic garnet composites comprise specimens with insignificant Ti and Na (LZGAR3) or significant Ti but insignificant Na (LZGAR4) contents.

The harzburgitic garnets (HZGAR1, 2) have raised Nd and Sr concentrations (Table 2), which belie their residual major element compositions. Furthermore, their Sm/Nd and ¹⁴³Nd/¹⁴⁴Nd ratios are lower, and ⁸⁷Sr/⁸⁶Sr ratios higher, than bulk Earth values for 1,180 Myr ago. These characteristics are similar to those for harzburgitic garnet inclusions in diamonds from Cretaceous kimberlites at Kimberley and Finsch, which yielded Rb-Sr and Sm-Nd model ages of 3.2–3.3 Gyr and required a metasomatically enriched (high Rb/Sr, low Sm/Nd) diamond precursor^{1,16}. Although this suggests a similar Archaean age and origin for the Premier harzburgitic diamonds, their garnet Sm/Nd and ¹⁴³Nd/¹⁴⁴Nd ratios (Fig. 2) are not sufficiently different from each other to yield an isochron age, or from bulk Earth values to yield a meaningful model age. On a Nd isotope evolution diagram (used to derive model ages; not shown), the harzburgitic garnet evolution curves are subparallel to the bulk Earth curve with intersection beyond, and hence model age greater than, the age of the Earth. Nevertheless, their antiquity is also manifest in the radiogenic ⁸⁷Sr/⁸⁶Sr ratios of the inclusion garnets (HZGAR1, 0.7043; HZGAR2, 0.7078) and corresponding unencapsulated macrocryst garnets (MGAR06, 0.7334; MGAR14, 0.7798), which are even higher than those of their Finsch and Kimberley counterparts¹. These garnets are essentially Rb-free and must incorporate radiogenic Sr by diffusive exchange with their Rb-enriched host rocks. The increase in Rb/Sr ratio produced by metasomatism would be variable on a local scale, accounting for the generation of more radiogenic Sr in the precursor of HZGAR2 relative to HZGAR1 during a potential 200–300 Myr interval between metasomatism and diamond crystallization¹. Thereafter, the ⁸⁷Sr/⁸⁶Sr ratios of inclusion and macrocryst garnets diverge owing to the isolation of inclusions from, and exposure of macrocrysts to, the progressively more radiogenic Sr generated by ⁸⁷Rb decay during storage in the continental mantle keel^{1,17}. Given that the mantle residence time of macrocrysts with such radiogenic Sr, sampled by the Premier kimberlite, is ~1 Gyr less than in the case of Cretaceous kimberlites, their host rocks must have had higher Rb/Sr ratios than at Finsch and Kimberley.

Two other macrocryst garnets with higher Ca contents approaching those of lherzolitic inclusions (Fig. 1) have similarly low ⁸⁷Sr/⁸⁶Sr ratios (MGAR01, 0.7021; MGAR12, 0.7027) indicating that they are roughly cogenetic and distinct from the macrocryst garnets already discussed. They define a two-point

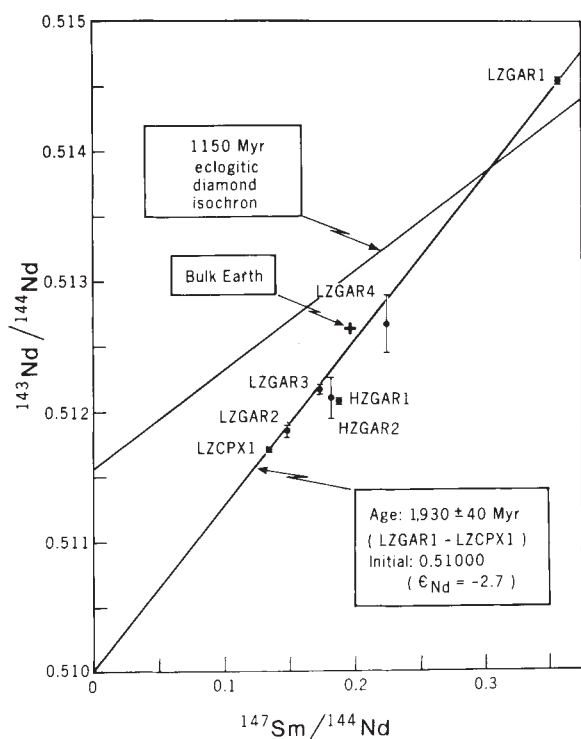


FIG. 2 Sm-Nd isochron diagram for peridotitic garnet and clinopyroxene inclusions in Premier diamonds (sample abbreviations as in Table 1). The age is calculated for the best-defined lherzolitic garnet-clinopyroxene pair (LZGAR1-LZCPX1), using $\lambda_{\text{Sm}} = 6.54 \times 10^{-12} \text{ yr}^{-1}$, with assigned error derived from maximum and minimum slopes allowed by $2\sigma_{\text{mean}}$ measurement errors. The 1,150 Myr eclogitic diamond isochron⁶ and bulk Earth value¹⁹ are shown for reference.

TABLE 2 Isotope composition of peridotitic garnet and clinopyroxene inclusions and macrocrysts from the Premier kimberlite

	n*	Wt(mg)	Rb	Sr	⁸⁷ Rb/ ⁸⁶ Sr	⁸⁷ Sr/ ⁸⁶ Sr	Sm	Nd	¹⁴⁷ Sm/ ¹⁴⁴ Nd	¹⁴³ Nd/ ¹⁴⁴ Nd	ε _{Nd}
Macrocryst grains											
MGAR01		2.65	0.0001	0.656	0.0005	0.70216 (4)	1.002	2.169	0.2794	0.51341 (4)	+2.5
MGAR06		5.13	0.0001	1.311	0.0003	0.77986 (2)	0.2632	1.027	0.1549	0.51267 (7)	+6.9
MGAR12		6.51	0.0004	0.390	0.0030	0.70279 (6)	1.753	1.706	0.6219	0.51609 (2)	+3.0
MGAR14		3.30	0.003	0.598	0.013	0.73340 (3)	0.9348	1.708	0.3309	0.51360 (19)	-1.5
Inclusions in diamonds											
HZGAR1	18	0.45	0.01	2.701	0.015	0.70431 (3)	2.233	7.187	0.1878	0.51208 (2)	
HZGAR2	3	0.12	—	3.22	—	0.70778 (6)	1.922	6.380	0.1821	0.51210 (16)	
LZGAR1	32	0.72	0.001	0.536	0.0034	0.70550 (10)	0.6741	1.141	0.3574	0.51454 (3)	
LZGAR2	20	0.78	0.002	1.317	0.0040	0.70378 (4)	0.5269	2.150	0.1482	0.51185 (5)	
LZGAR3	29	0.83	0.002	0.824	0.0056	0.70450 (7)	0.2683	0.9353	0.1734	0.51217 (4)	
LZGAR4	30	0.67	0.001	0.749	0.0043	0.70592 (6)	0.3785	1.019	0.2246	0.51267 (22)	
LZCPX1	32	0.43	0.008	93.96	0.0002	0.70379 (1)	0.9805	4.414	0.1342	0.51171 (2)	
LZCPX2	23	0.38	0.062	111.3	0.0016	0.70369 (2)	0.4565	—	—	—	

Small-sample preparation and analysis were performed in the Radiogenic Isotope Facility at the University of Cape Town using techniques refined at the Massachusetts Institute of Technology^{1,2,6}. Isotope ratios were measured on a VG Sector 7-collector mass spectrometer in multidynamic mode, with ion yields up to 5% for Sr and Nd metal ions. Concentrations are in p.p.m. (μg g⁻¹), after correction for minimum applicable blanks: Rb, 3 pg; Sr, 4 pg; Sm, 0.7 pg; Nd, 1.6 pg. Estimated analytical uncertainties in concentrations (ignoring small-sample weighing errors) are indicated by the number of significant digits. Present day ⁸⁷Sr/⁸⁶Sr and ¹⁴³Nd/¹⁴⁴Nd ratios are normalized to 0.71022 for NBS-987 and 0.51264 for BCR-1 (equivalent to 0.51184 for La Jolla), respectively. In-run precision is indicated by the errors (2σ_{mean}) in the least significant digits given in parentheses. Macrocryst initial ⁸⁷Sr/⁸⁶Sr (not listed; age corrections same order of magnitude as in-run precision) and ¹⁴³Nd/¹⁴⁴Nd (expressed as ε_{Nd} = {(¹⁴³Nd/¹⁴⁴Nd)_{sample} / (¹⁴³Nd/¹⁴⁴Nd)_{bulk Earth}} - 1) × 10⁴) ratios were calculated for a kimberlite emplacement age of 1,180 Myr.

*Number of diamonds from which composite was derived (sample name abbreviations as in Table 1).

Sm-Nd isochron age of 1,190 ± 20 Myr, which is equivalent to the megacryst ages on which the kimberlite emplacement age of 1,180 ± 30 Myr is based^{8,9}. The initial ⁸⁷Sr/⁸⁶Sr and ¹⁴³Nd/¹⁴⁴Nd (ε_{Nd} = +2; see Table 2) ratios are indicative of a trace-element-depleted (low Rb/Sr, high Sm/Nd) source similar to those for the host kimberlite (⁸⁷Sr/⁸⁶Sr = 0.7025; ε_{Nd} = +4; S.H.R. unpublished data) and eclogitic diamonds (⁸⁷Sr/⁸⁶Sr = 0.7022; ε_{Nd} = +8; ref. 6).

The best-defined lherzolitic garnet-clinopyroxene pair (LZGAR1-LZCPX1) yields a two-point Sm-Nd isochron age of 1,930 ± 40 Myr (Fig. 2). Other lherzolitic garnets (LZGAR2, 3, 4) with intermediate Sm/Nd ratios also lie on the isochron, within the larger analytical errors. The corresponding initial ¹⁴³Nd/¹⁴⁴Nd ratio (ε_{Nd} = -3) indicates a mildly trace-element-enriched (low Sm/Nd, high Rb/Sr) precursor. This is also reflected in the initial ⁸⁷Sr/⁸⁶Sr ratios of the garnets and clinopyroxenes (0.7037-0.7058) which are significantly above that of the bulk Earth at the time (~0.7023). Ideally, these essentially Rb-free inclusions should all have the same ⁸⁷Sr/⁸⁶Sr ratio if they are cogenetic^{2,6}. Although one garnet composite (LZGAR2) has the same ⁸⁷Sr/⁸⁶Sr ratio (0.7037) as the two clinopyroxene composites, the other three (LZGAR1, 3, 4), with lower Sr contents (Table 2), indicate a contribution of more radiogenic Sr from one or more grains in each composite. The formation of lherzolitic diamonds in pre-existing lithosphere may involve mixing of asthenospheric and lithospheric components, as for eclogitic

diamonds^{2,3,6}. The consequences for initial ¹⁴³Nd/¹⁴⁴Nd ratios are difficult to access and we concede that the isochron relationship in Fig. 2 might partially result from such mixing and not have strict age significance. In the worst case, the lherzolitic diamond age would still probably lie between those of harzburgitic and eclogitic diamonds from this locality.

Our preferred lherzolitic diamond formation age (1930 Myr) and precursor isotope signature (mildly enriched) contrast with those of the first-generation harzburgitic diamonds (~3,200 Myr; strongly enriched) formed during Kaapvaal shield stabilization⁷ and the third-generation eclogitic diamonds (~1,150 Myr; strongly depleted) formed just before kimberlite emplacement⁶. This preferred lherzolitic diamond age is only ~100 Myr less than the ages of the basic layered rocks (~2,060 Myr) and granites (~2,050 Myr) of the adjacent Bushveld Complex^{5,18}. The Rustenburg Layered Suite represents one of the largest mantle melting events beneath the Kaapvaal craton during the Proterozoic. Its overall composition is basaltic¹³ and its mantle residua would be less depleted in basaltic components (that is, more lherzolitic) than those of komatiites. A potential compositional link thus exists between such residua and lherzolitic diamond precursors, given metasomatic introduction of diamond-forming components after emplacement of the Bushveld Complex. This temporal, spatial and compositional link is analogous to that between komatiitic magmatism and subsequent harzburgitic diamond formation in the Archaean¹. □

Received 14 June; accepted 20 September 1993.

- Richardson, S. H., Gurney, J. J., Erlank, A. J. & Harris, J. W. *Nature* **310**, 198-202 (1984).
- Richardson, S. H., Erlank, A. J., Harris, J. W. & Hart, S. R. *Nature* **346**, 54-56 (1990).
- Smith, C. B. et al. *Geochim. cosmochim. Acta* **55**, 2579-2590 (1991).
- Gurney, J. J. & Switzer, G. S. *Contr. Miner. Petrol.* **39**, 103-116 (1973).
- Hamilton, J. J. *Petrol.* **18**, 24-52 (1977).
- Richardson, S. H. *Nature* **322**, 623-626 (1986).
- de Wit, M. J. et al. *Nature* **357**, 553-562 (1992).
- Smith, C. B. thesis, Univ. Witwatersrand (1983).
- Jones, R. A. thesis, Univ. Leeds (1984).
- Burgess, R., Turner, G., Laurenzi, M. & Harris, J. W. *Earth planet. Sci. Lett.* **94**, 22-28 (1989).
- Phillips, D., Onstott, T. C. & Harris, J. W. *Nature* **340**, 460-462 (1989).
- Aillson, H. L., Burger, A. J. & van Zyl, C. *Earth planet. Sci. Lett.* **3**, 161-166 (1967).
- Eales, H. V. et al. *J. Afr. Earth Sci.* **16**, 121-142 (1993).

- Gurney, J. J., Harris, J. W., Rickard, R. S. & Moore, R. O. *Trans. geol. Soc. S. Afr.* **88**, 301-310 (1985).
- Deines, P., Harris, J. W., Spear, P. M. & Gurney, J. J. *Geochim. cosmochim. Acta* **53**, 1367-1378 (1989).
- Shimizu, N. & Richardson, S. H. *Geochim. cosmochim. Acta* **51**, 755-758 (1987).
- Richardson, S. H. in *Continental Mantle* (ed. Menzies, M. A.) 55-65 (Clarendon, Oxford, 1990).
- Walraven, F., Armstrong, R. A. & Kruger, F. J. *Tectonophysics* **171**, 23-48 (1990).
- Wasserburg, G. J., Jacobsen, S. B., DePaolo, D. J., McCulloch, M. T. & Wen, T. *Geochim. cosmochim. Acta* **45**, 2311-2323 (1981).

ACKNOWLEDGEMENTS. We acknowledge De Beers Consolidated Mines, Kimberley, for diamond specimens and the Foundation for Research Development, Pretoria, for financial support. We thank P. Bartlett, C. R. Clement, the late A. J. Erlank, S. R. Hart, J. B. Hawthorne, I. M. Laithwaite, R. S. Rickard, J. A. Robey, N. Wilson-Harris and the perennial diamond sorters, V. Anderson, G. Parker and E. van Blerk for their contributions.

Intrinsic density-dependent regulation of vole populations

Richard S. Ostfeld*, Charles D. Canham*
& Stephen R. Pugh*†‡

* Institute of Ecosystem Studies, Box AB, Millbrook,
New York 12545, USA

† College of General Studies, Boston University, Boston,
Massachusetts 02215, USA

‡ Present address: University of New Hampshire at Manchester,
Manchester, New Hampshire 03102, USA

CONSIDERABLE controversy exists over the role of density-dependent processes in controlling animal population size. In populations that fluctuate cyclically or erratically, for example many voles and insects^{1,2}, theory predicts that either density-dependence is weak^{1,3}, or that density-dependent responses lag behind density⁴⁻⁶. One key mechanism for lagged density-dependence is a delay in regeneration of food resources following heavy exploitation. Here we show that meadow vole (*Microtus pennsylvanicus*) populations respond immediately to high density by reducing breeding effort and hence population growth, disproving the hypothesis that density-dependence is weak. In addition, vole populations do not show a delay in growth following marked reduction in plant biomass (their source of food and cover). We conclude that intrinsic density-dependence processes tend to stabilize vole populations, and that cyclic dynamics are not caused by lagged effects of resource exploitation.

Vole populations often fluctuate cyclically, with 3–5 years between peaks^{7,8}. To mimic different points in the natural cycle, we established three replicates each of meadow vole populations held chronically at low, medium and high density in nine field enclosures. Populations averaged about 70, 180 and 380 voles per hectare, respectively (Fig. 1), maintained at targeted densities by removing subadults (the primary age of dispersal^{9,10}). Enclosures were closely juxtaposed and therefore experienced similar abiotic and biotic conditions. Avian and mammalian predators were not excluded by the fences.

Correlational studies of vole population dynamics have shown that neither mortality rates nor emigration rates are density-dependent⁹⁻¹¹. Instead, typical populations appear to be characterized by catastrophic declines after the achievement of a peak. Peaks are usually preceded by atypically high rates of winter breeding, which are critical to vole population dynamics^{7,12}. The

determinants of winter breeding, however, have until now not been established^{11,12}.

In our study, population density determined the rate of autumn, winter and spring breeding in both 1990–1991 and 1991–1992 (Fig. 2a, b). Females and males on high-density grids became non-reproductive 2–4 weeks earlier in the autumn, and reproductive four weeks later in the spring than voles on medium-density grids. On low-density grids, although there was a reduction in the proportion of voles breeding between October and May, roughly half bred throughout this period, resulting in continuous recruitment (Fig. 2c). High-density treatments resulted in 2–3 fewer generations per year than medium-density treatments, and up to 5 fewer generations per year than low-density treatments. As vole generation times are short (3–4 weeks^{13,14}; typically 5–10 generations per year) and litter size is large (~4–5; refs 13,14), the loss of several generations per year would sharply curtail population growth rate.

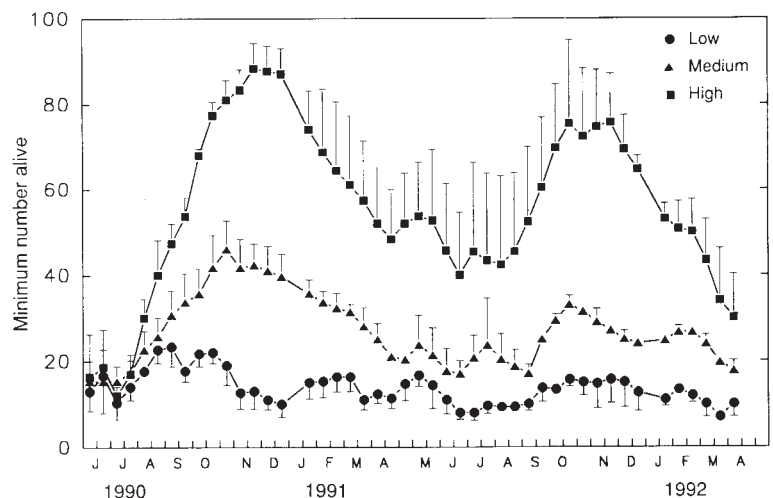
Vole density had little effect on juvenile survival probabilities as determined by repeated measures analysis of variance in both breeding and non-breeding seasons of both years of the study. Similarly, analysis of variance on median mass at sexual maturity for both sexes and all seasons only revealed a significant effect of density for females during the 1991–1992 non-breeding season ($F=6.02$; d.f. = 2; $P=0.038$; ranking, low < medium < high). Therefore the number, but not the quality, of offspring appears to be strongly density-dependent.

Female (but not male) voles under all density treatments curtailed breeding during an unusual summer drought during June and July 1991 (Fig. 2a, b). Thus, voles gave a density-independent response to an unpredictable change in resource availability, namely the drying of their principal food, grasses and forbs. Despite an apparent consensus that both density-dependent and density-independent processes affect population dynamic patterns^{3,15-20}, clear examples of both processes acting on the same population are few.

Above-ground plant biomass in spring was significantly reduced in high- and medium-density enclosures compared with low-density enclosures, but recovered by late summer or autumn; moreover, food quality was reduced by the elimination of some preferred plant species and the reduction of others (R.S.O., manuscript submitted). Thus, the effects of voles on plants seemed likely to cause lagged density-dependence. Models incorporating lags of about nine months produce cyclic dynamics with a period and magnitude virtually identical to those of microtine populations^{4,21}.

Lagged density-dependence has been detected in time series

FIG. 1 Densities of the experimental vole populations expressed as minimum number alive per 0.16 hectare enclosure. Symbols show means (± 1 s.e.) of the 3 replicates of low-, medium-, and high-density treatments, which were arranged in a randomized block design, in an old field in south-eastern New York state. The data set consists of 11,282 capture records of 3,094 individuals in the enclosures. The densities maintained represent the range of naturally occurring densities for unenclosed populations of this species^{8,11}. Low- and medium-density enclosures were maintained at desired levels by removing selected subadults (body mass, 20–30 g) during normal biweekly live trapping. Subadults, independent of their mothers, were targeted to minimize disruption of social structure of the populations, and to simulate dispersal that occurs primarily in subadults⁹⁻¹¹. Individuals were not removed from designated high-density enclosures. Sex ratios of removed individuals were equalized within and among enclosures. Owing to high reproductive rates in low-density enclosures, and to the necessity of preserving realistic genetic diversity and sex ratios, we were unable to maintain densities more typical of nadirs in cyclic vole populations (≤ 50 animals per hectare¹¹).



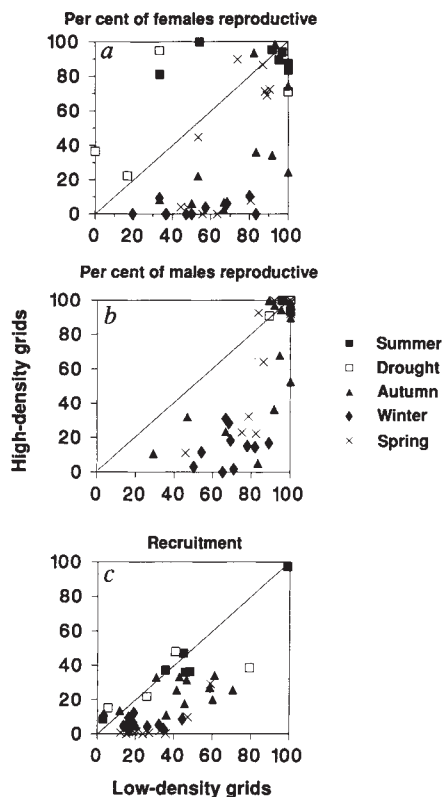


FIG. 2 Phase diagrams comparing high-density and low-density treatments with respect to: a, per cent of females in breeding condition; b, per cent of males in breeding condition; and c, recruits as a per cent of the total adult population. Females were considered to be in breeding condition if the vaginal orifice was open²⁹, and males were considered to be in breeding condition if their testes were descended into the scrotum¹⁴. Each data point is the combination of the mean of three replicates of high density and the mean of three replicates of low density for a given trapping session. Trapping sessions are distinguished by season: June–August, summer; September–November, autumn; December–February, winter; March–May, spring. In addition, the four summer data points taken during the 1991 drought (in which mean monthly rainfall was <40% of the 1951–1980 average for this site) are distinguished. Diagonal line indicates an equivalent proportion of individuals breeding (or recruiting) at high and low density; points below the diagonal line indicate density-dependence.

of insects^{5,22}, and of voles²³, but the mechanisms causing delayed density-dependence are poorly understood and controversial. Four primary hypotheses have been put forward to explain delayed density-dependence in voles: behavioural polymorphism¹¹, maternal effects²⁴, predation⁸, and herbivore–resource^{8,11,21}. Recent experimental and theoretical studies^{24–27} substantially weaken the first two hypotheses. Moreover, the lack of a strong density effect on either survival or maturation rates of juveniles does not support the prediction of the maternal effects hypothesis, that high population density causes stress to females, reducing the quality of their offspring. This leaves predator–prey and herbivore–resource dynamics as the most plausible mechanisms for lagged density-dependence. We therefore tested whether the depression of resource quantity and quality, resulting from high vole density, caused a time lag in the ability of vole populations to recover.

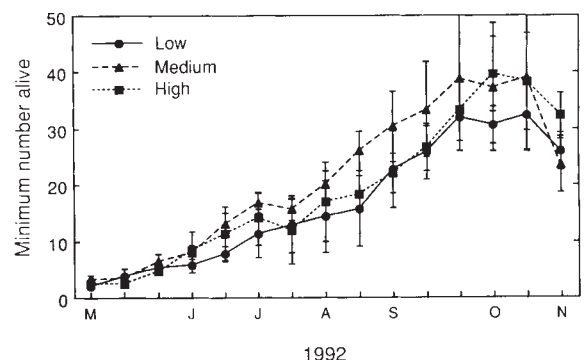
In April 1992, after 20 months of chronically different densities among treatments, all voles from the experimental enclosures were removed and one week later two pairs of voles, caught 2 km away, were introduced into each enclosure. One week later, missing voles were replaced to equalize sex and age composition

among enclosures. The timing of the removal and introduction simulated the annual 'spring decline'¹¹ characteristic of North American voles. Populations were then allowed to grow freely to November 1992, after which the study ended.

A time lag in the growth of populations introduced into enclosures that were previously high-density was not observed; dynamics were similar to those in enclosures previously of low- or medium-density. Although there were no statistically significant differences in density or growth rate among the experimental treatments, voles introduced into formerly medium-density treatments maintained the highest density levels on average (Fig. 3). This result suggests that a moderate degree of grazing by voles may stimulate plant growth, whereas high and low grazing intensity are less stimulating.

Our demonstration of instantaneous density-dependent reproduction supports the hypothesis that intrinsic (demographic and behavioural) features of vole populations tend to be stabilizing, and that extrinsic features (predators and disease agents) may be necessary to cause the dramatic population fluctuations often observed^{25,27,28}. Our experimental results indicate that a consumer–resource lag is unlikely to contribute to cyclic popula-

FIG. 3 Population growth trajectories of the vole populations inhabiting enclosures that had previously experienced high, medium or low population density for 20 months. Symbols represent the means (minimum number alive per 0.16 hectare) of the three replicates of each prior-density treatment. Repeated measures analysis of variance revealed no significant differences in density among prior-density treatments during any trapping session.



tion dynamics in this system, and are consistent with analyses²³ supporting a key role for delayed density-dependence resulting from predator-prey dynamics. □

Received 15 April; accepted 7 September 1993.

- Andrewartha, H. G. & Birch, L. C. *The Distribution and Abundance of Animals* (University of Chicago, 1954).
- Elton, C. W. *Voles, Mice and Lemmings* (Clarendon, Oxford, 1942).
- Strong, D. R. *Trends Ecol. Evol.* **1**, 39–42 (1986).
- May, R. M. in *Theoretical Ecology: Principles and Applications* (ed. May, R. M.) 4–25 (Saunders, Philadelphia, 1976).
- Turchin, P. *Nature* **344**, 660–663 (1990).
- Hassell, M. P. *Trends Ecol. Evol.* **1**, 90–93 (1986).
- Cockburn, A. *Social Behaviour in Fluctuating Populations* (Croom Helm, London, 1988).
- Hansson, L. & Henttonen, H. *Trends Ecol. Evol.* **3**, 195–200 (1988).
- Gaines, M. S. & McClenaghan, L. R. *A. Rev. ecol. Syst.* **11**, 163–196 (1980).
- Lidicker, W. Z. in *Biology of New World Microtus* (ed. Tamarin, R. H.) 420–454 (Am. Soc. Mammal. Spec. Publ., no. 8, 1985).
- Krebs, C. J. & Myers, J. H. *Adv. ecol. Res.* **8**, 267–399 (1974).
- Schaffer, W. M. & Tamarin, R. H. *Evolution* **27**, 111–124 (1973).
- Hasler, J. F. *The Biologist* **57**, 52–86 (1975).

- Keller, B. L. in *Biology of New World Microtus* (ed. Tamarin, R. H.) 725–778 (Am. Soc. Mammal. Spec. Publ., no. 8, 1985).
- Hassell, M. P. *J. Anim. Ecol.* **56**, 705–713 (1987).
- Dempster, J. P. & Pollard, E. *Oikos* **46**, 413–416 (1986).
- Brown, M. W. *Ecology* **70**, 776–779 (1989).
- Stiling, P. *Ecology* **70**, 779–783 (1989).
- Hanski, I. *Phil. Trans. R. Soc. B* **330**, 141–150 (1990).
- Gaston, K. J. & Lawton, J. H. *Oecologia* **74**, 404–410 (1987).
- Ostfeld, R. S. in *Wildlife 2001: Populations* (eds McCullough, D. R. & Barrett, R. H.) 851–863 (Elsevier, London, 1992).
- Woiwod, I. P. & Hanski, I. *J. Anim. Ecol.* **61**, 619–629 (1992).
- Hanski, I., Turchin, P., Korpimäki, E. & Henttonen, H. *Nature* **364**, 232–234 (1993).
- Mihok, S. & Boonstra, R. *Can. J. Zool.* **70**, 1561–1566 (1992).
- Stenseth, N. C. *Theor. Populat. Biol.* **29**, 365–385 (1986).
- Chitty, D. *Can. J. Zool.* **65**, 2555–2566 (1987).
- Akçakaya, H. R. *Ecol. Monogr.* **62**, 119–142 (1992).
- Hanski, I., Hansson, L. & Henttonen, H. *J. Anim. Ecol.* **60**, 353–367 (1991).
- McCravy, K. W. & Rose, R. K. *J. Mammal.* **73**, 151–159 (1992).

ACKNOWLEDGEMENTS. We thank C. Borg, D. Braun, G. Capreol, J. Jimenez, K. Kinder, P. Nielson-Murphy, S. Plambeck, K. Price, H. Rolland Fischer, J. Schnurr and A.-O. Wilhelm for help in the field, and M. Shachak, J. Van Buskirk and J. Wolff for comments. S.R.P. was supported by a Summer Research Fellowship from the Mary Flagler Cary Charitable Trust. Funding was from Central Hudson Gas and Electric Corporation, Empire State Electric Energy Research Corporation, General Reinsurance Corporation, and the Mary Flagler Cary Charitable Trust.

New discoveries of *Australopithecus* at Maka in Ethiopia

Tim D. White*, Gen Suwa†, William K. Hart‡, Robert C. Walter§, Giday WoldeGabriel||, Jean de Heinzelin¶, J. Desmond Clark#, Berhane Asfaw** & Elisabeth Vrba††

* Laboratory for Human Evolutionary Studies, Department of Anthropology, University of California, Berkeley, California 94720, USA

† Department of Anthropology, University of Tokyo, Bunkyo-ku, Hongo, Tokyo 113, Japan

‡ Department of Geology, Miami University, Oxford, Ohio 45056, USA

§ Geochronology Center, Institute of Human Origins, 2453 Ridge Road, Berkeley, California 94709, USA

|| EES-1/D462, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

¶ Institut Royal des Sciences Naturelles de Belgique, 29 Vautier Street, 1040 Brussels, Belgium

Department of Anthropology, University of California, Berkeley, California 94720, USA

** Paleoanthropology Laboratory, PO Box 5717, Addis Ababa, Ethiopia

†† Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06511, USA

THE taxonomy of *Australopithecus afarensis*, the oldest known hominid species, has been a matter of debate since its description in 1978 (ref. 1). Some authorities regard all specimens assigned to *A. afarensis* as belonging to a single taxon^{2–4} whereas others regard the Tanzanian and Ethiopian specimens as each representing a different species^{5,6}. Further controversy surrounds the issues of sexual dimorphism and locomotion among these hominids. Resolution of these problems would shed light on hominid phylogeny in general and on the ancestry of later *Australopithecus* and *Homo*. Fossils discovered in the Afar of Ethiopia in 1990 constitute the first major addition to the 3–4 million year (Myr) hominid record since the 1970s. We report here the discovery of new fossils from Maka, dated to 3.4 Myr ago, which provide powerful support for the interpretation of *A. afarensis* as a single, ecologically diverse, sexually dimorphic, bipedal Pliocene primate species whose known range encompassed Ethiopia and Tanzania.

The Rift Valley Research Mission in Ethiopia recovered the Middle Pleistocene Bodo cranium^{7–9}. Our team subsequently provided radiometric dates and the first Pliocene hominids from the Middle Awash, including a proximal femur from Maka^{10,11}.

The Pliocene sedimentary record along the eastern margin of the Middle Awash begins with thick deposits indicating semi-continuous sedimentation in a large lake¹². As the lake's influence diminished about 3.85 Myr 'pedi-alluvial' aggradation began. This represents the recurrent interference between lateral pedimentation from wadis or seasonal rivers on the east side of the basin, and alluviation related to a major river in the basin centre. The geography, stratigraphy, and geochronology of the area are summarized in Fig. 1.

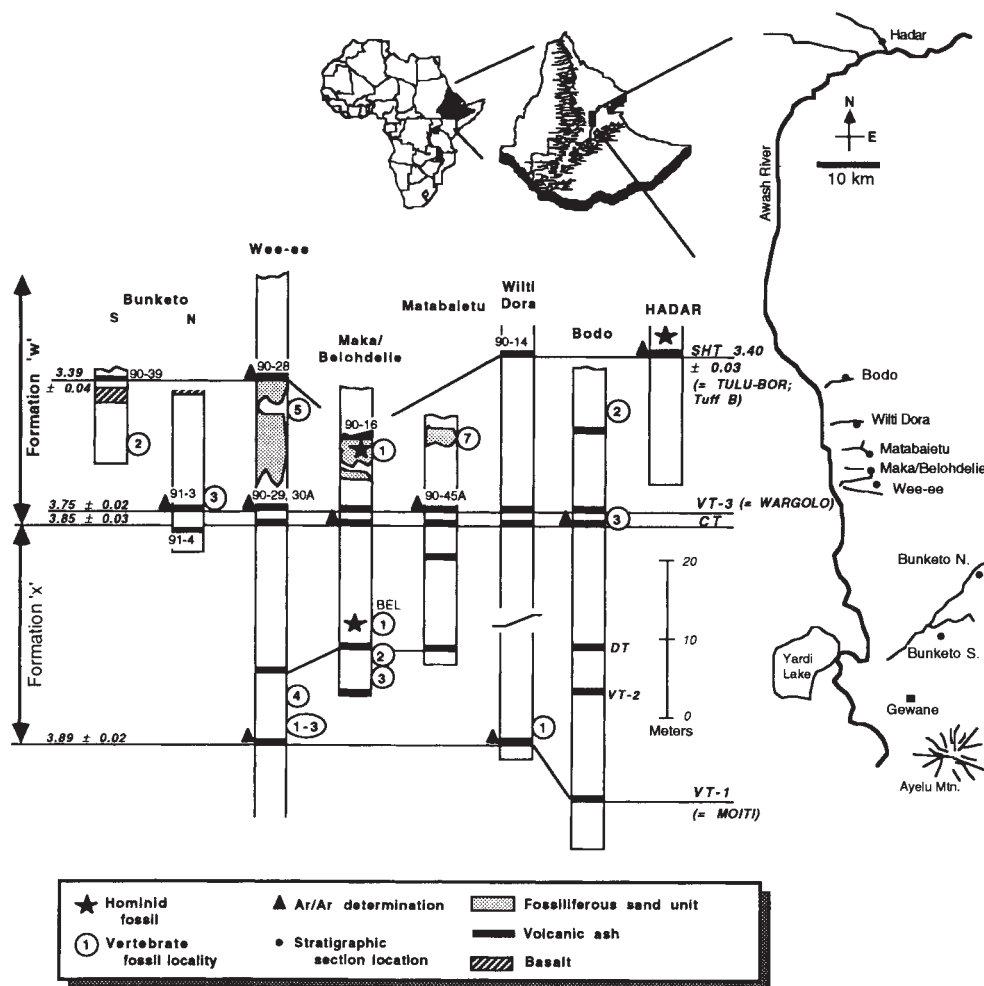
The description of *A. afarensis*² provoked debate over systematics and locomotion in the earliest known hominids. The paratype series of this species was assembled from two locations, Hadar in Ethiopia and Laetoli in Tanzania. Nomenclature^{5,13}, the number of species included in *A. afarensis*^{6,14–19}, and the validity of pooling Laetoli and Hadar specimens^{6,20} were questioned. Some viewed the species as a dedicated biped^{21–23} and others viewed it as a locomotor 'missing link' with retention of strong quadrupedal and/or climbing adaptations^{24–26}. The 1990 Maka discoveries provide crucial evidence on these issues.

The new Maka hominid finds include a right mandible with M_{1–3} (MAK-VP-1/2); a left humerus (1/3); a right M₂ (1/4); a left edentulous mandible (1/6); a nearly complete mandible with dentition (1/12); a left M₁ (1/13); a left mandible ramus (1/83); and a left proximal ulna (1/111). All specimens were surface finds as depicted in Fig. 1. Tephrostratigraphic (Fig. 1; Table 1), radiometric (Fig. 1; Table 2), and biochronological results indicate a reliable age estimate of about 3.4 Myr for the Maka hominids. Discovery of the SHT/Tulu-Bor beta tuff in the mid-Pliocene of the Middle Awash is the first tephrostratigraphic link between the hominid-bearing deposits at Hadar and those of the Middle Awash. This shows that the Hadar Formation and the Middle Awash Pliocene sections overlap in time and that schemes attempting chronologically to sever these two sets of deposits²⁷ are therefore incorrect.

The MAK-VP-1/4 molar (RM₂; 16.2 mm length, 13.8 mm breadth) is larger than any previously known *A. afarensis* homologue except A.L. 188–1 (ref. 28). The new MAK-VP-1/13 molar (mesiodistal crown diameter of 10.9 mm) now defines the small end of the *A. afarensis* range. Ulnar fragment MAK-VP-1/111 matches the A.L. 288–1 'Lucy' specimen in size and confirms the presence of contemporary small individuals. The new Maka sample thus shows dramatic intraspecific variation in tooth and body size in a temporally circumscribed sample, supporting similar indications at Laetoli³ and Hadar Locality 333 (ref. 29).

Specimen MAK-VP-1/3 (Fig. 2a, top) is the first fairly complete humerus from a large, presumably adult male individual of this antiquity. Like fragmentary Hadar counterparts (A.L. 333–87, –107, –109 (ref. 29)), it is very robust for its length, with midshaft cortical thickness of up to 8 mm (medial surface).

FIG. 1 Map and columns showing the new radiometric determinations, geometric relationships, and litho-stratigraphic/tephrochemical correlations of volcanic horizons in the western Afar. Middle Awash areas are identified with Afar names for the catchments studied. Initial mapping of Pliocene deposits³⁴ was followed by Taieb's naming the 'Formation D'Urugus' in the Bunketo drainage³⁵. The Rift Valley Research Mission in Ethiopia (RVRME) placed these rocks in a Pliocene 'Sagantole Formation'^{7,36}. This stratigraphic nomenclature is not used here because its underpinning descriptions and definitions of beds and marker horizons lack precision and validity. The Cindery Tuff (CT) is a widespread marker that is used here as a formational boundary. Below the CT is a predominantly lacustrine sequence we informally name formation 'x.' Above is an increasingly pedimental formation 'w.' We previously correlated VT-1 with the Moiti tuff of the Turkana Basin³⁷, a correlation recently confirmed^{38,39}. This tuff occurs in the Main Ethiopian Rift⁴⁰, the Omo River Valley^{39,41}, and Deep Sea Drilling Project (DSDP) cores from the Gulf of Aden and Indian Ocean³⁸. We date VT-1 at 3.89 ± 0.02 Myr (Table 1). Our Sibabi marker tuff date of 4.26 Myr (Table 1) shows that the Moiti is not the oldest tuff in this area as opposed to ref. 27. Initial dating placed the Cindery Tuff between 3.8 and 4.0 Myr⁴². New single grain dates for CT plagioclases range from 3.03 to 4.67 Myr. The weighted mean age of 21 dates is 3.85 ± 0.08 Myr. The BEL-VP-1/1 hominid frontal¹¹ is thus between 3.89 and 3.86 Myr. Tuff VT-3 is usually found within 2 m above the CT. We chemically correlate it from Bunketo to Bodo (Table 2), date it at 3.75 ± 0.02 Myr, and confirm that it is the same as the Wargolo Tuff of Northern Kenya and tuffs from Gulf of Aden DSDP cores^{27,38}. The proximity of the fossiliferous Maka sands to tuffs below is not a reliable age indicator because of the intervening erosional unconformity¹⁰. Our 1990 fieldwork at nearby Matabailetu and Wee-ee established additional fossil localities in similar sand and gravel beds. Two additional volcanic horizons were discovered. The first, MA 90-28 at Wee-ee, chemically correlates with tuffs at Bunketo and Wilti Dora (Table 1). Eight anorthoclase grains from this unit yield a mean date of 3.39 ± 0.04 Myr. This is precisely identical to the age of the SHT tuff at the base of the Hadar Formation⁴³, with which it chemically correlates (Table 2). A tuff in Gulf of Aden core 231 and the Tulu-Bor beta tuff in the Turkana Basin³⁸ also correlate. Biochronological considerations



support this correlation. Tuff MA 90-16 is embedded within the hominid-bearing MAK-VP-1 sands and gravels. Radiometric dating of feldspars from this ash demonstrated pervasive Miocene contamination; thus no direct age estimate was possible. This tuff yields two distinct glass populations; a dominant heterogeneous high Fe population, and a subordinate low Fe population that is chemically similar to the Tulu-Bor beta (although Ca more similar to Tulu-Bor alpha) tuff of the Turkana Basin³⁸ (Table 2). Neither MA90-16 nor the underlying unit yielding the fossils at Maka can be traced physically to fossiliferous sands in the Wee-ee area some 1.5–2.0 km away. But the striking sedimentological and stratigraphic similarities, the occurrence of Tulu Bor/SHT type glass in MA90-16, and the matching abundant and diverse vertebrate fossils from both areas (and from the Laetoli Beds and Hadar's Sidi Hakoma Member) indicate that the SHT/Tulu-Bor beta age of about 3.4 Myr⁴³ is a reliable age estimate for the new hominids from Maka.

There is a large, laterally flaring deltoid tuberosity, an extremely well developed lateral supracondylar ridge that imparts a laterally convex distal shaft margin, a retroflexed shaft, and a deep pit for the insertion of pectoralis major. The contrast between this and the A.L. 288-1 humerus (Fig. 2a, bottom) parallels human sex-related morphological differences and reinforces interpretation of these remains as dimorphic but conspecific. The Maka femur¹⁰, the Laetoli footprints⁵, and the Hadar pelvis, knee and foot²¹ indicate that the *A. afarensis* hindlimb was specialized for terrestrial bipedality. Compared to the straight-shafted humeri of modern African apes, the contemporary Maka humerus is very robust. This discovery lends strong support to the inference that *A. afarensis* retained a powerful upper limb, but an upper limb that lacked the key arboreal adaptation of great length²².

Four 1990 mandibular specimens conform metrically and morphologically to patterns established by the Hadar hominids²⁹. Specimen MAK-VP-1/12 (Fig. 2d) is a mandible with I_2 - M_3 , RP_3 - M_3 , and portions of both condyles. Assembled from 109 fragments, this specimen is the most complete mid-Pliocene hominid mandible known. In dental and osseous size, it is centred in the *A. afarensis* range (Fig. 2b). Characters used to attribute it to that species include lateral corpus contour hollowing, large canines and incisors, inferiorly placed mental foramina, asymmetrical third premolars, a C/P_3 diastema, an $I_2 + C$ to P_3 occlusal step-down, straight, anteriorly convergent tooth rows, prominent lingual molar cusps, and a 'serrate' molar root pattern²⁹. The strong buccal face slope of the P_4 , the bilobed lateral molar profiles, and the large trigonid breadth in the M_2 are also features shared with *A. afarensis*. The P_3 is bicuspid,

TABLE 1 Single crystal laser-fusion $^{40}\text{Ar}/^{39}\text{Ar}$ results for Middle Awash tephra

L number	Mat.	Ca/K	% $^{40}\text{Ar}^*$	Age (Myr)	$\pm(1 \text{ s.d.})$	L number	Mat.	Ca/K	% $^{40}\text{Ar}^*$	Age (Myr)	$\pm(1 \text{ s.d.})$
Within Maka sands, Wee-ee (MA90-28)						Cindery Tuff, Gamedah (CT-G and Bodo (A81-1)					
4984-01	A	0.1127	100.9	3.486	0.164	2638-01	P	13.7700	61.5	3.536	0.219
4984-02	A	0.0855	94.3	3.440	0.270	2638-02	P	12.3220	85.2	4.150	0.148
4984-03	A	0.0576	93.5	3.328	0.070	2638-03	P	6.974	88.7	3.907	0.106
4984-04	A	0.0985	98.6	3.371	0.304	2638-04	P	6.874	21.1	3.035	0.385
4984-05	A	0.0786	98.6	3.555	0.165	2638-05	P	6.595	83.6	3.439	0.278
4994-02	A	0.0779	94.9	3.461	0.222	1851-01	P	4.2741	49.3	4.666	0.262
4994-04	A	0.1097	96.6	3.318	0.142	1851-02	P	4.8218	41.4	3.085	0.231
4994-06	A	0.1028	97.7	3.476	0.136	1851-03	P	3.5679	45.3	3.407	0.215
Avg.		0.0904	96.9	3.390	0.036	1851-04	P	5.5648	53.1	3.851	0.235
s.d.		0.0188	2.5			1851-05	P	7.4236	56.4	4.054	0.248
4994-05†	A	0.0787	99.9	4.087	0.146	1851-06	P	4.6227	54.9	3.918	0.231
4994-01†	A	0.1049	98.4	5.413	0.356	1851-08	P	2.0384	30.3	3.807	0.426
VT-3, Matabaietu (MA90-45A) and Wee-ee (MA90-29 and -30)						1851-09	P	6.4598	52.0	3.727	0.301
5013-02	S	0.0300	86.7	3.650	0.073	1851-10*	P	2.1685	51.9	3.977	0.178
5013-03	S	0.0129	95.2	3.662	0.034	1851-11*	P	5.6358	47.1	3.368	0.162
5013-04	S	0.0239	96.3	3.669	0.136	1851-12*	P	6.6289	65.1	3.546	0.168
5013-05	S	0.0181	92.8	3.670	0.045	1851-13*	P	6.6280	72.1	4.152	0.194
5006-01	S	0.0127	99.9	3.735	0.033	1851-14*	P	4.9971	45.1	4.143	0.194
5006-02	S	0.0361	96.6	3.612	0.061	1851-15*	P	6.7645	54.6	4.376	0.170
5006-03	S	0.0269	100.2	3.885	0.059	1851-16*	P	4.5418	54.0	3.708	0.233
5006-04	S	0.0123	99.0	3.823	0.071	Avg.		6.1337	55.6	3.847	0.083
5007-01	S	0.0102	99.8	3.763	0.045	s.d.		2.8254	17.1		
5007-02	S	0.0165	97.1	3.825	0.041	VT-1, Wilti Dora (A81-36)					
5007-03	S	0.0233	100.1	3.863	0.047	2662-02	S	0.020	17.2	3.898	0.238
5007-04	S	0.0712	97.4	3.698	0.040	2662-03	S	0.027	91.3	3.879	0.073
5007-05	A	0.1157	97.7	3.811	0.034	2662-04	S	0.015	88.9	3.901	0.050
Avg.		0.0274	96.8	3.748	0.023	2662-06	S	0.000	89.7	3.999	0.120
s.d.		0.0277	3.7			2662-07	A	0.247	97.6	4.065	0.276
5013-01†	S	0.0173	98.4	5.661	0.044	2662-08	S	0.016	92.3	3.786	0.086
5006-05†	S	0.0307	81.8	5.842	0.048	2662-09	S	0.129	74.9	3.868	0.137
Marker Tuff, Sibabi (A81-56)						2662-10	S	0.012	94.4	3.905	0.080
2656-04	S	0.033	90.2	4.240	0.031	Avg.		0.0583	80.8	3.889	0.023
2656-05	S	0.026	41.3	4.332	0.095	s.d.		0.0863	26.5		
2656-06	S	0.060	92.3	4.265	0.048	2662-05†	S	0.008	98.9	16.363	0.079
2656-09	S	0.066	42.0	4.212	0.068						
2656-10	S	0.038	93.0	4.302	0.039						
Avg.		0.0446	71.8	4.263	0.019						
s.d.		0.0172	27.5								
2656-02†	P	13.677	100.9	4.724	0.285						
2656-03†	S	0.083	26.9	4.792	0.150						
2656-08†	S	0.023	94.2	4.769	0.071						

New radiometric dates for Middle Awash tephra based on single crystal laser fusion (SCLF) analysis. L number is laboratory run number. Ca/K is derived by multiplying the measured $^{37}\text{Ar}/^{39}\text{Ar}$ ratio by 1.96. The bold mean ages are weighted means and uncertainties ($\pm$, 1 s.d.) computed using an inverse variance weighting factor²⁸. Individual grain errors reflect errors in the measurement of J (the neutron fluence parameter) and in the determination of Ar isotopic ratios, which in turn propagate errors in discrimination (D) and errors in Ar beam intensities from sample and blank. Data in italics (†) are excluded from the mean age. Feldspars were irradiated for 0.5 h at 8 MW at the Omega West research reactor of the Los Alamos National Laboratory, which has a fast neutron fluence of $5.7 \times 10^{13} \text{ n cm}^{-2} \text{ s}^{-1}$. Cadmium shielding was used to reduce thermal neutron production of ^{40}Ar . After irradiation, the samples were transferred to a copper holder and loaded onto the extraction line for overnight bakeout at about 200 °C. Samples were fused with an 8 W argon-ion laser. Abundances of the Ar isotopes were measured on a MAP-215 noble gas spectrometer fitted with a Balzers electron multiplier operating at a gain of about 20,000. Typical system blank volumes of ^{40}Ar , ^{39}Ar , ^{38}Ar , ^{37}Ar , and ^{36}Ar are 4, 2, 0.08, 0.3 and 0.3×10^{-17} moles, respectively. Sanidine from the Fish Canyon Tuff was used as the neutron fluence monitor, which has a reference age of 27.84 Myr. Constants used in this study follow ref. 33. S, sanidine; P, plagioclase; A, anorthoclase (based on Ca/K). Detailed results for individual grain analyses are available from R.C.W. by request.

with a developed mesiolingual corner, as in specimens from A.L. 333. This mandible is metrically and morphologically a good match for the *A. afarensis* holotype L.H.-4 from Laetoli, Tanzania (Fig. 2c). It displays degeneration of the articular surfaces of both condyles, a pathology unexpected for an individual with relatively unworn third molars. It constitutes the earliest evidence of bilateral temporomandibular joint disease among hominids.

The new Maka hominids are accompanied by fauna indicative of an environment broadly comparable to that represented by

the later Denen Dora fauna at Hadar. This setting was ecologically intermediate between the contemporary open, dry Laetoli environment and the more closed, mesic Hadar SH Member environment. The broad habitat range inferred for these early hominids predicts that additional remains will be found across a wide geographic range in Africa. Some have stressed the antiquity of the Tanzanian Laetoli hominids relative to those from Hadar, emphasizing the intervening time gap²⁷. Others have suggested gradual evolutionary change^{30,31} and/or taxonomic distinctiveness³² based on the ecological, geographic and tempo-

TABLE 2 Compositional data for Middle Awash tephra*

	Loc.	†	SiO ₂	TiO ₂	Al ₂ O ₃	Fe ₂ O ₃	MnO	MgO	CaO	Na ₂ O	K ₂ O	TOTAL	Zn	Zr	Nb	Ba	Sr	Rb	Y
(a) SHT/BETA-TULU BOR/TUFF B-BETA																			
SHT‡	HAD	DCP	72.37	0.16	12.02	1.66	0.05	0.09	0.33	4.83	1.93	93.44	80	333	81	187	10	118	68
MA90-14	WIL	DCP	72.48	0.15	12.05	1.59	0.06	0.12	0.32	4.37	3.22	94.36	77	331	82	173	10	125	69
MA90-14	WIL	MP	70.89	0.12	11.75	1.43	0.07	0.05	0.29	4.49	3.55	92.64							
MA90-28	WEE	DCP	72.60	0.14	12.00	1.52	0.05	0.07	0.32	4.22	3.51	94.43	79	344	81	212	11	135	70
MA90-28	WEE	MP	71.05	0.12	11.41	1.49	0.03	0.02	0.28	4.37	3.87	92.64							
MA90-39	BUN	MP	71.33	0.10	11.40	1.44	0.06	0.03	0.28	4.61	2.29	91.53							
(b) MAK SAND TUFF																			
MA90-16	MAK	MP																	
LoFe			71.50	0.10	12.07	1.51	0.03	0.03	0.48	3.91	4.17	93.80							
HiFe			69.21	0.14	11.85	2.11	0.05	0.05	0.73	3.90	3.70	91.73							
(c) VT-3/WARGOLO																			
VT-3‡	MAK	DCP	74.02	0.14	11.15	2.37	0.07	0.04	0.19	3.67	3.74	95.39	180	787	124	83	3	117	123
MA90-29	WEE	DCP	73.07	0.14	10.81	2.40	0.07	0.06	0.16	3.81	3.73	94.25	174	756	127	73	5	115	119
MA90-29	WEE	MP	71.72	0.11	10.39	2.15	0.02	0.02	0.13	3.73	3.80	92.07							
MA90-30A	WEE	DCP	73.48	0.14	10.67	2.41	0.07	0.05	0.17	3.85	3.80	94.63	195	758	129	71	4	110	121
MA90-30A	WEE	MP	72.14	0.12	10.35	2.18	0.07	0.02	0.16	4.12	4.34	93.50							
MA90-45A	MAT	DCP	73.00	0.13	10.69	2.37	0.06	0.06	0.16	3.80	3.98	94.24	175	741	127	74	4	120	120
MA90-45A	MAT	MP	70.96	0.12	10.28	2.19	0.06	0.00	0.18	4.12	4.21	92.12							
TW91-3	BUN	DCP	73.66	0.14	10.69	2.34	0.07	0.03	0.22	4.15	3.08	94.38	182	744	130	68	2	108	124
TW91-3	BUN	MP	71.53	0.12	10.29	2.23	0.04	0.01	0.17	4.34	3.19	91.92							

* Major element oxides in weight per cent as measured with total Fe expressed as Fe₂O₃; trace elements in parts per million (p.p.m.).

† Glass shards from all 1990 and 1991 tephra samples were analysed for major element oxides by an automated Cameca Electron Microprobe (MP) operated at 12–20 nA beam current. Standards used for calibration include a natural glass and a set of feldspars, pyroxenes, and amphiboles. When pure glass separates (>99.5%) could be obtained and only a single glass population was identified by MP analyses, major and trace element concentrations were analysed by Direct Current Argon Plasma (DCP) Spectrometry. DCP major element concentrations were determined by comparison of unknown intensities to calibration curves based on six internationally recognized rock standards. DCP trace element concentrations were determined by the method of standard addition using a multi-element spike solution designed specifically for silicic compositions.

‡ Samples of SHT and VT-3 from the originally defined localities were analysed for major and trace element concentrations by DCP in the same batches as the 1990–91 tephra to aid comparisons and correlations.

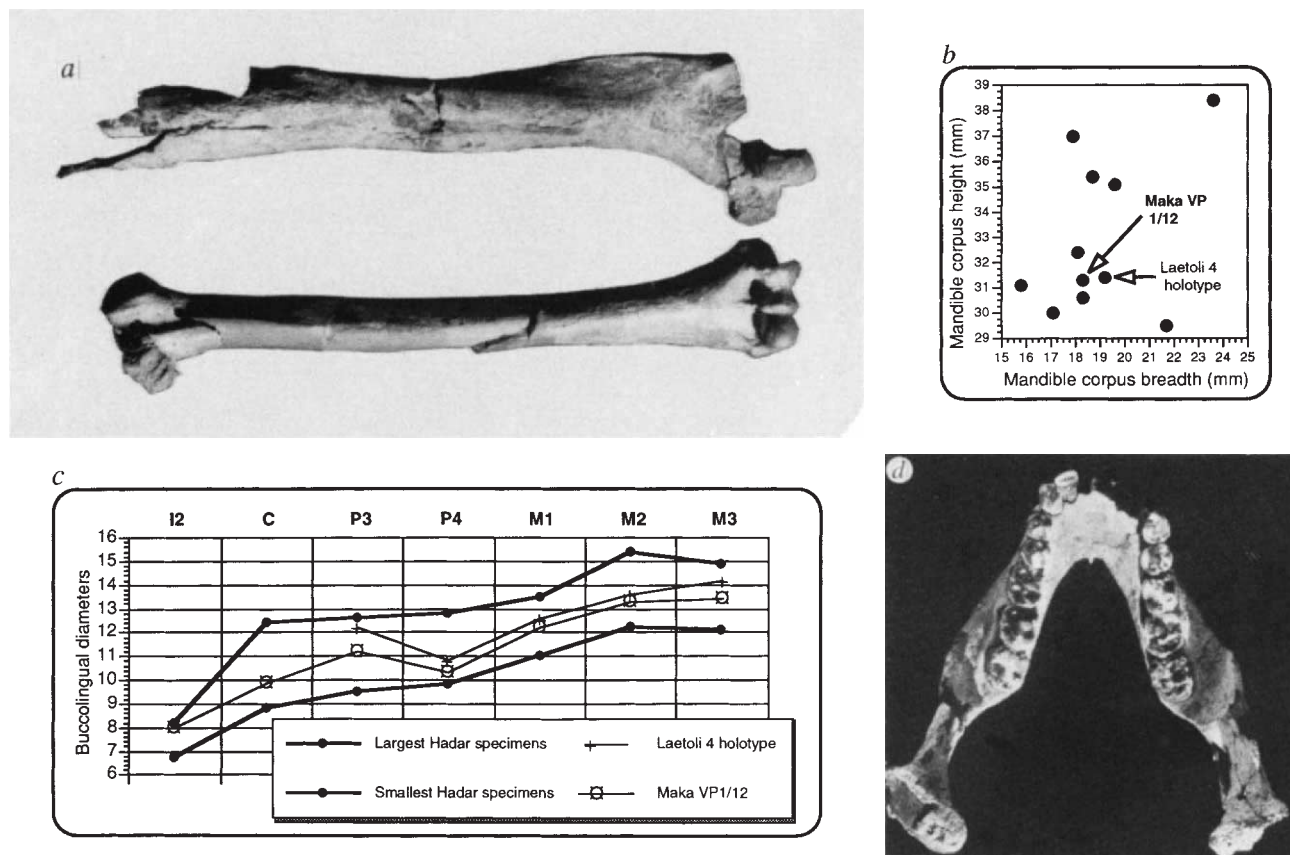


FIG. 2 a, The new Makha humerus compared to the smaller A.L. 288-1 ('Lucy') humerus from Hadar, shown 40% natural size. b, Mandible corpus dimensions for Hadar *A. afarensis*²⁹, the Laetoli holotype L.H.-4, and the new Makha mandible. c Buccolingual breadth dimensions of mandibular teeth for the largest and smallest Hadar specimens at

each tooth position (bold lines) compared to homologous measures on the *A. afarensis* holotype and the new Makha mandible. d, Photograph of the MAK-VP-1/12 specimen, the most complete mandible of *A. afarensis*, shown 40% natural size.

ral separation. The MAK-VP-1/12 mandible's strong resemblance to the nearly contemporary Laetoli holotype of *A. afarensis* supports arguments for pooling these Ethiopian and Tanzanian fossils and provides strong evidence of stasis within this species². □

Received 11 June; accepted 31 August 1993.

- Johanson, D. C., White, T. D. & Coppens, Y. *Kirtlandia* **28**, 1–14 (1978).
- Johanson, D. C. & White, T. D. *Science* **202**, 321–330 (1979).
- White, T. D. in *Ancestors: The Hard Evidence* (ed. Delson, E.) 139–152 (Liss, New York, 1985).
- Boaz, N. T. *Yearb. Phys. Anthropol.* **31**, 85–113 (1988).
- Leakey, M. D. *Phil. Trans. R. Soc. Lond. B* **292**, 95–10 (1981).
- Tuttle, R. H. *Rev. Anthropol.* **17**, 391–426 (1988).
- Kalib, J. E. et al. *Nature* **298**, 17–25 (1982).
- Kalib, J. E. et al. *Nature* **298**, 25–29 (1982).
- Conroy, G. C., Jolly, C. J., Cramer, D. & Kalib, J. *Nature* **275**, 67–70 (1978).
- Clark, J. D. et al. *Nature* **307**, 423–428 (1984).
- Asfaw, B. *J. hum. Evol.* **16**, 611–624 (1987).
- Adamson, D. A. & Williams, M. A. *J. hum. Evol.* **16**, 597–610 (1987).
- Tobias, P. V. *Palaeont. afr.* **23**, 1–17 (1980).
- Leakey, R. E. *Am. Sci.* **64**, 174–178 (1976).
- Coppens, Y. *Le Singe, l'Afrique et l'Homme* (Fayard, Paris, 1983).
- Olson, T. R. in *Aspects of Human Evolution* (ed. Stringer, C.) 99–128 (Taylor and Francis, London, 1981).
- Senut, B. *Bull. Mem. Soc. Anthropol. Paris* **10**, 325–334 (1983).
- Zihlman, A. in *Hominid Evolution: Past, Present and Future* (ed. Tobias, P. V.) 213–220 (Liss, New York, 1981).
- Falk, D. *Braindance* (Holt, New York, 1992).
- Mayr, E. in *A History of American Physical Anthropology 1930–1980* (ed. Spencer, F.) 231–237 (Academic, New York, 1982).
- Latimer, B. & Lovejoy, C. O. *Am. J. phys. Anthropol.* **78**, 369–386 (1989).
- Latimer, B. in *Origine(s) de la Bipédie chez les Hominidés* (eds Coppens, Y. & Senut, B.) 169–176 (Editions du CNRS, Paris, 1991).
- Lovejoy, C. O. *Sci. Am.* **256** (11), 118–125 (1988).
- Stern, J. T. & Susman, R. L. *Am. J. phys. Anthropol.* **60**, 279–317 (1983).
- Rose, M. D. in *Origine(s) de la Bipédie chez les Hominidés* (eds Coppens, Y. & Senut, B.) 37–48 (Editions du CNRS, Paris, 1991).
- Susman, R. L. & Stern, J. T. in *Origine(s) de la Bipédie chez les Hominidés* (eds Coppens, Y. & Senut, B.) 121–131 (Editions du CNRS, Paris, 1991).
- Haileab, B. & Brown, F. H. *J. hum. Evol.* **22**, 453–468 (1992).
- Deino, A., Tauxe, L., Monaghan, M. & Drake, R. J. *Geol.* **98**, 567–587 (1990).
- Johanson, D. C. et al. *Am. J. phys. Anthropol.* **57**, 373–719 (1982).
- Cronin, J. E., Boaz, N. T., Stringer, C. B. & Rak, Y. *Nature* **292**, 113–122 (1981).
- Leonard, W. R. & Hegmon, M. *Am. J. phys. Anthropol.* **73**, 41–63 (1987).
- Day, M. H. *Nature* **300**, 574 (1982).
- Walter, R. C., Manega, P. C., Hay, R. L., Drake, R. E. & Curtis, G. H. *Nature* **354**, 145–149 (1991).
- Gortani, M. & Biachi, A. *Missione Geologica Dell'azienda Generale Italiana Petroli (AGIP) Nella Danalia Meridionale e Sugli Altipiani Hararini (1936–1938)* (Accademia Nazionale dei Lincei, Rome, 1973).
- Taieb, M. thesis, Univ. Paris (1974).
- Kalib, J. E., Oswald, E. B., Mebrate, A., Tebedge, S. & Jolly, C. J. *Newsl. Stratigr.* **11**, 95–127 (1982).
- Walter, R. C., WoldeGabriel, G. & Westgate, J. *Geol. Soc. Am. Abstr. Progr.* **17**, 743 (1985).
- Brown, F. H., Sarna-Wojcicki, A. M., Meyer, C. E. & Haileab, B. *Quat. Int.* **13/14**, 55–67 (1992).
- Hart, W. K., Walter, R. C. & WoldeGabriel, G. *Quat. Int.* **13/14**, 77–86 (1992).
- WoldeGabriel, G. & Aronson, J. L. *Geology* **15**, 432–433 (1987).
- WoldeGabriel, G., Walter, R. C., Aronson, J. L. & Hart, W. K. *Quat. Int.* **13/14**, 69–76 (1992).
- Hall, C. M., Walter, R. C., Westgate, J. A. & York, D. *Nature* **308**, 26–31 (1984).
- Walter, R. C. & Aronson, J. L. *J. hum. Evol.* **25**, 229–240 (1993).

ACKNOWLEDGEMENTS. We thank the Anthropology and Archaeometry programs of the National Science Foundation, the National Geographic Society, the Japan Society for the Promotion of Science, the University of California Collaborative Research Program of the Institute of Geophysics and Planetary Physics at Los Alamos National Laboratory, the Afar people, the Ethiopian Ministry of Culture and Sports Affairs, the National Museum of Ethiopia, A. Ademassu, Y. H-Selassie, B. Latimer, K. Schick, S. Yirga, G. Richards, A. Angela, A. Asfaw, E. Zerfu, A. Negash, Z. Assefa, G. Hundie and C. Tilahun for making the 1990 Middle Awash research possible.

Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome

S. Brenner*, G. Elgar, R. Sandford, A. Macrae, B. Venkatesh† & S. Aparicio

MRC Molecular Genetics Unit and Department of Medicine, University of Cambridge Clinical School, Addenbrooke's Hospital, Cambridge CB2 2QQ, UK

CLONING and sequencing techniques now allow us to characterize genes directly instead of having to deduce their properties from their effects. This new genetics reaches its apotheosis in the plan to obtain the complete DNA sequence of the human genome, but this is far beyond the capacity of present sequencing methods. Small 'model' genomes, such as those of *Escherichia coli* (4.7 megabases (Mb)¹ and yeast (14 Mb)², or even those of *Caenorhabditis elegans* (100 Mb) and *Drosophila* (165 Mb), are better scaled to existing technology. The yeast genome will contain genes with functions common to all eukaryotic cells, and those of simple multicellular organisms may throw light on the genetic specification of more complex functions. However, vertebrates differ in their morphology and development, so the ideal model would be a vertebrate genome of minimum size and complexity but with maximum homology to the human genome. Here we report the characterization of the small genome (400 Mb) of the tetraodontoid fish, *Fugu rubripes*³. A random sequencing approach supported by gene probing shows that the haploid genome contains 400 Mb of DNA, of which more than 90% is unique. This genome is 7.5 times smaller than the human genome and because it has a similar gene repertoire it is the best model genome for the discovery of human genes.

* Address for correspondence: Department of Medicine, University of Cambridge Clinical School, Addenbrooke's Hospital, Cambridge CB2 2QQ, UK.

† On leave from the Institute of Molecular and Cell Biology, National University of Singapore, Singapore.

It has been known for some time that the genomes of tetraodontoid fish are particularly small. Hinegardner found haploid DNA contents ranging from 0.4 pg (380 Mb) in *Tetraodon fluviatilis* to 0.5 pg (480 Mb) in *Spheroides (Arothron) maculatus*⁴, about one-eighth to one-sixth of the size of the human genome. Measurements of the reassociation kinetics of the DNA of another tetraodontoid fish, *Arothron diadematus*⁵, revealed that 13% of the DNA consisted of sequences with varying degrees of repetition, with the remaining 87% probably being unique and having a complexity 80 times that of *E. coli* DNA. This sequence component corresponds to 410 Mb and the total genome to 470 Mb, based on length measurement of the *E. coli* genome⁶, which agrees with measurements for *Spheroides*^{6,9}. These methods of measuring genome complexity are not always reliable, however, so we sought an independent method to validate the small genome size reported in these fish. We found that sequencing random genome fragments not only provides direct information about the abundance and composition of different classes of DNA, including known genes, but also allows the genome size to be estimated from the fraction of identified genes.

Tables 1 and 2 summarize the results of sequencing 596 clones generated by sonication of genomic DNA. Repeat sequences and ribosomal gene sequences were found in 108 clones and comprise 9,762 base pairs (bp) (7.6%) of the total sequence. The most abundant minisatellite, with a repeat unit of 118 bp, represents almost 2% of the total genome and is highly clustered (Fig. 1). We found no significant homology to Alu or LINE sequences or other abundant mammalian dispersed repeats and our results indicate that, with the exception of some microsatellites, most or all of the high-copy-number repetitive DNA is clustered in the genome. Forty four sequences consist almost entirely of clustered repeats. This includes all minisatellites, RNA genes and some microsatellites. The remaining 552 sequences were translated into amino-acid sequences in all six reading frames and searched for homologies against the SWISSPROT protein database. Ten sequence fragments were found with very strong similarity to corresponding regions of known human and other mammalian genes. A summary of these results can be found in Table 2. An example demonstrates not only the good correspondence between the amino-acid sequences of the exons, but the intron

boundaries, as defined by potential splice acceptor and donor sites, are also in exactly the same positions in both sequences.

Table 2 also shows that the ten identified genes provide a total of 1,011 bp of coding sequence, which is 0.791% of the total nucleotides sequenced. We can use this information on coding sequences to estimate the size of the *Fugu* genome. To obtain the

TABLE 1 Analysis of repeat sequences from randomly sheared fragments of the *Fugu* genome

Type	Repeat unit (bp)	No. of clones	Total length of sequence (bp)
Minisatellite			
N118	118	13	2,543
NRF1	complex	4	755
NRF2	33	2	225
P1d7	23	1	368
P2a8	30	1	240
P2e8	12/13	1	275
P4e3	21	1	63
P4f9	29	1	174
P4g8	40	1	240
2130	12/18	1	113
281	13	1	91
P2a11	33	1	132
P3h6	79	1	272
1173	62	1	186
2g4	35/36	1	143
2h7	10/11	1	102
	Total:	32	5,922
Microsatellite			
A_nT_m (n or $m > 4$)	complex	2	20
Telomeric	AGGGTT	6	442
Penta-	AGAGG	5	125
Tetra-	CAGA	4	76
	(A/G)GAT	4	148
	(A/G)CAG	4	100
Tri-	CAT-ATG	3	238
	ACT-AGT	3	135
Di-	(GT) $_n$ $n \geq 9$	2	1,050
	(CT) $_n$ $n \geq 10$	2	134
	(AG) $_n$ $n \geq 10$	2	122
	(AT) $_n$ $n \geq 10$	2	28
	(GC) $_n$ $n \geq 10$	2	1
Mono-	(A-T) $_n$ $n \geq 10$	1	193
	(G-C) $_n$ $n \geq 10$	1	34
	Total:	72	2,845
Ribosomal			
	5S + spacer	1	234
	18S + 28S + spacer	3	761
	Total:	4	995
Total non-coding			108
			9,762

Repeat sequences and ribosomal gene sequences were found in 108 clones and represent 9,762 bp (7.6%) of the total sequence. The minisatellite repeats comprise 4.6% of the sequence, and one minisatellite sequence, with a repeat of 118 bp, accounts for 2,543 bp of sequence and therefore represents almost 2% of the total genome. None of the other minisatellite repeats have any resemblance to known interspersed or clustered repeats in other vertebrates. The number of sequences with different repeat units is probably not exhausted as there are a large number of single recurrences. Microsatellite or simple sequence repeats comprise 2.2% of the total bases sequenced, and one of these, the telomeric repeat (AGGGTT) $_n$, with 442 bp of sequence, represents 0.35% of the genome. GT-AC repeats are the most abundant dinucleotide repeat with 30 examples, followed by CT-AG and AT repeats, each with 4 examples. These ratios, 30:4:4, may be compared with the relative abundances of 73:19:16 of these sequences in the human sequence database¹¹. The distributions of the lengths of the repeats are similar to those recorded in the human genome¹¹. The absolute frequency of GT-AT repeats in the human genome has been experimentally determined to be 1 per 30 kb on human chromosome 16, similar to a value of 1 per 28 kb computed from GenBank¹². The corresponding value for the *Fugu* genome is one every 4,261 bp. Finally, the 18S and 28S rRNA genes account for 761 bp of sequence and 0.6% of the genome; the 5S genes are less abundant with a content of about 0.2%.

comparable figure for the human genome, we first constructed a non-redundant coding sequence database, called RATMAN (see legend to Table 2). This contains 3,090,858 bp of coding sequence: a search of random sequence from the human genome of 3,000 Mb for similarities against SWISSPROT would be expected to identify 0.103% as coding sequence. On the assumption that the *Fugu* and human genomes have the same number of genes, we can conclude that the pufferfish genome, where the fraction of known coding sequences is 0.791%, is $0.791/0.103 = 7.68$ times smaller than the human genome, that is, 390 Mb. An independent estimate of genome size was made by hybridizing unamplified libraries with single-copy-gene probes. Table 3 shows that, on average, one genome is contained in 24,625 phages. With an average insert size of 16.4 kilobases (kb), this predicts a genome size of 404 Mb assuming that all sequences are equally represented in the library. Both new estimates accord with the smaller genome size of 380 Mb found in *Tetradon*

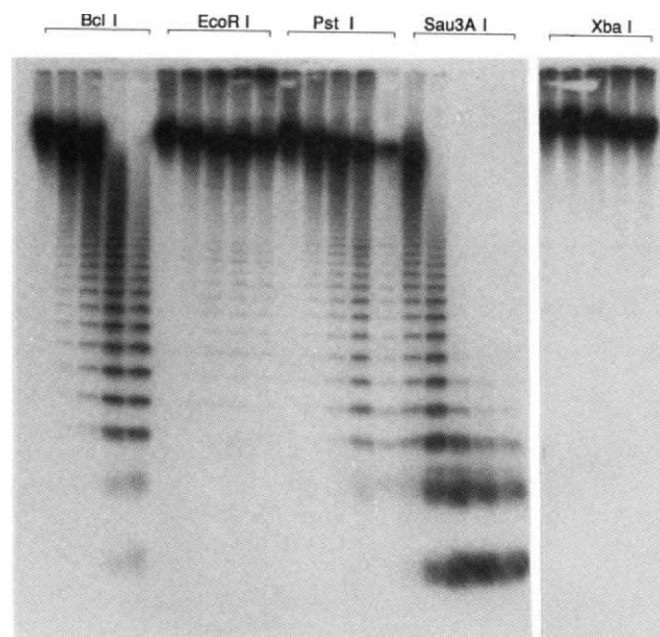


FIG. 1 *Fugu* genomic DNA was digested to increasing degrees (left to right) with *BclI*, *EcoRI*, *PstI*, *Sau3AI* or *XbaI* and probed with a PCR-generated ³²P-dCTP labelled 118-bp repeat. *Sau3AI* sites are present in all of the repeat units sequenced so far, followed in order of frequency by *BclI*, *PstI*, *EcoRI* and finally *XbaI*, which is not present in any. This distribution of sites is reflected in the hybridization profiles of these enzymes, which demonstrate the highly clustered nature of this repeat.

```

ple51      GTTTTCAGGTGATCATGTTGAATTTACCTCTGTTTGGAGAACTGTATATCCTGACCAA
ssp2b32    GTTTTCAGATGATCATGTTGCATTACGCTCTGTTTGCAGAAAATGTCTCTGACCAA
ssp2d91.p  -----CCAA
ssp2d93.p  GTTTTCAGATGATCATGTTGCATTACCTCTGTTTGGAAATAATGTATCTCTGACCAA
consensus  GTTTTCAGATGATCATGTTGCATTACCTCTGTTTGGANAGKWTGTNTCTCTGACCAA
Human rpt 1 GTTTTCAGATGATCATGTTGCATTACCTCTGTTTGGAAATAATGTATCTCTGACCAA
Human rpt 2 GTTTTCAGGTGATCATGTTGCATTACCTCTGTTTGGAGAACTGTATCTCTGACCAA
ple51      AAGTGATGGTTTCCCCACGAGAAAACGTCAGAAAACGTCATAATGTGACCGCAGCATGA
ssp2b32    A-GTGATGGTTTCCCCACGAGAAAACGTCAGAAAACGTCATAATGTGACCGCAGCATGA
ssp2d91.p  AAGTGATGGTTTCCCCACGAGAAAACGTCAGAAAACGTCATAATGTGACCGCAGCATGA
ssp2d93.p  AAGGGATGGTTTCCCCACGAGAAAACGTC-----
consensus  AAGTGATGGTTTCCCCACGAGAAAACGTCAGAAAACGTCATAATGTGACCGCAGCATGA
Human rpt 1 AAGTGATGGTTTCCCTCCAGAACACGTCAGAAAACGTCATAATGTGACCGCAGCATGA
Human rpt 2 AAGTGATGGTTTCCCCACGAGAAAACGTCAGAAAACGTCATAATGTGACCGCAGCATGA

```

Alignment of members of the 118-bp-repeat family in *Fugu* represent the more common polymorphisms. Also shown are clones generated from genomic DNA by PCR. A consensus sequence is provided for the *Fugu* sequences. Thirteen clones contained different permutations of this sequence showing that it is tandemly repeated in the genome; a consensus sequence of the repeat unit reconstructed from the sequence data is shown. PCR and sequencing reveal that the same sequence exists in the human genome but probably at a much lower frequency. The EMBL accession numbers for sequences mentioned in this paper are X74416 and X74417.

TABLE 2 Analysis of coding sequences from randomly sheared fragments of the *Fugu* genome

Gene	Coding	Intron	Other	Total
Human fibroblast growth factor receptor	58	42	—	100
Rat tropomyosin-2 α -chain, brain	135	57	81	273
Human cAMP-dependent protein kinase A	72	14 + 253	—	339
Human placental protein PHPS1-2	60	286	—	346
Mouse inositol 1,4,5-trisphosphate-binding	120	156	—	276
Rat cytochrome P450 IIB1, also IID1	92	108	—	200
Human synaptotagmin fragment (P65)	75	24	221	320
Human kringle (hepatocyte growth factor)	173	55	—	228
Human integrin (β 3 subunit)*	163	8 + 73	—	244
Rat B-myc transforming protein	63	11	9	83
Total:	1,011	997	311	2,409

DNA was sheared by sonication, the ends repaired and the blunt-end fragments cloned in the *EcoRV* site of Bluescript. Two different bacteria were used for cloning: about 1/3 of the library was cloned in *E. coli* DH5, and the remaining 2/3 in *E. coli* Sure (Stratagene), which does not select against methylated DNA. As there was no difference between the two sets, the results have been pooled. The clones were sequenced, mostly by radioactive methods, using the dideoxy chain termination method¹⁴; 596 sequences were obtained, providing a total of 127,831 bp of sequence, with an average of 214 bp per clone. The sequences were matched against themselves and then against the EMBL (rel.31) and SWISSPROT (rel.21) databases using BLASTX and BLASTN¹⁵ with default parameters to detect coding sequences and rRNA respectively. Dinucleotide repeats and minisatellites were found by clustering the sequences into related families using ICAtool¹⁶; single copy repeats were detected with the REPEATS programme from the UWGCG¹⁷ suite. The RATMAN database was constructed from the same version of SWISSPROT (release 21.0) used for the sequence searches. This database includes a single example of each mammalian coding sequence in SWISSPROT and is a list of all known genes that have either been found or could reasonably be expected to be found in the human genome. There is a total of 2,576 genes, of which 1,759 were human, 169 were bovine, 302 were mouse, and 346 were rat. Where several examples exist, precedence for inclusion was in the order human, bovine, mouse, rat. These sequences contain a total of 1,032,286 codons, or 3,096,858 nucleotides; the average length of a coding sequence is 1.2 kb.

*Comparison of *Fugu* β 3 integrin with the corresponding region of human platelet glycoprotein IIIa gene¹³.

```
Fugu  ..ttcag]a IQDKQALFGFKNVLSLSTEQVARPTTEVEKQQTSHNRDAPEGGFDAVMQAVVCK.. [gtagg
      . . . . .
Human ..cccag]t MKTTCLPMFGYKHVLTLDQVTRFNEEVKKQSVSRNRDAPEGGFDAIMQATVCD.. [gtgag
```

species^{5,8}. The close correspondence of the estimate based on random sequencing with those based on physical estimates validates the assumption of a similar gene repertoire.

All of the experiments point to a small genome size in *Fugu* of 400 Mb, 7.5 times smaller than the human genome. As the

repetitive sequences comprise 7.4% of the genome of 400 Mb, all of the functional genes must be accommodated in 370 Mb. We note that the ratio of distances between GT-AC repeats in the human and fish genomes is $30/4.26 = 7.05$, close to the ratio of genome sizes. This means that both have about the same number of repeats (100,000), which might indicate that the genome expansion in mammals occurred between the repeats. But as there is only 30% positional conservation of these repeats between different mammalian genomes¹⁰, it is unlikely that all the repeats are ancient. The compactness of the fish genome can be explained by the relative paucity of 'junk' sequences as compared with mammalian genomes. There is very little repetitive DNA in the *Fugu* genome, and it occurs largely in clusters. There are no dispersed repeats or pseudogenes, elements that are major components of mammalian genomes. We also have evidence (unpublished results) that the introns are very much smaller than their human equivalents: for example, three-quarters of *Fugu* introns are small (66–120 bp, modal value 80 bp). Initial sequencing of contiguous DNA confirms the expected high gene density.

Teleost fish have all the specialized functions of higher vertebrates, so most of the genes responsible for complex functions in higher vertebrates already exist in the fish genome. Sequencing part or even all of this compact genome, with its small introns and high relative content of exon information, would be an effective way of discovering vertebrate genes and provide relevant access to the human genome. We are seeking to explore this approach further by sequencing complex genes such as dystrophin, by exhaustively enumerating members of gene families and by studying contiguous genes to see how far short-range linkage is conserved between fish and mammal. In addition, the comparative function of sequences from this model are being tested in transgenic experiments. □

TABLE 3 Probing of a *Fugu* genome library with single gene probes

Probe	DSO library 3.5×10^4	GE library 1×10^5	DSN library 8.8×10^4
THA5	2		
PAHD1	1		
InsD4	2	8	7
PC3	1	5	
C12.Ex.Ec			3
Hepsin			2
Actin	3		
D2Rint		4	
Dystr8		4	
Dystr12		4	
MPO			0
GabaR			2
B7Xh0.Ec			3
Tyr			2
MicA			3
Positives	9	25	22
Phages	1.75×10^5	5×10^5	7.04×10^5
Total of 56 found in $1.379 \times 10^6 = 1$ in 24,625 phages			

Fugu DNA was partially digested with *Sau3A*1 and a 16–20 kb fraction isolated by electrophoresis in agarose. This was cloned into *Bam*H1-cut λ 2001 (ref. 18) and phages with inserts were selected by plating on strain CQ6 (*E. coli* C(P2)). The average size of inserts was later determined by restriction mapping of 70 independent phages to be 16.4 kb. Plaques from the 3 independent non-amplified libraries were probed with a variety of sequences isolated from other experiments. Probes were labelled with ³²P-dCTP using random hexamers or polymerase chain reaction (PCR). Filter hybridization and washing followed ref. 19. Autoradiographs were obtained after exposure to X-ray film at -70°C for 24–48 hours with image-intensifying screens. All positive plaques were confirmed either by secondary hybridization or by PCR with sequence-specific primers.

Received 28 April; accepted 8 September 1993.

1. Daniels, D. L., Plunkett, G., Burland, V. & Blattner, F. R. *Science* **257**, 771–778 (1992).
2. Oliver, S. G. *et al. Nature* **357**, 38–46 (1992).
3. Sulston, J. *et al. Nature* **356**, 37–41 (1992).
4. Ajioka, J. W. *et al. Chromosoma* **100**, 459–509 (1991).
5. *Pufferfishes Available in Japan* (Chuou-Houki, Tokyo, 1984).
6. Hinegardner, R. *Am. Nat.* **102**, 517–523 (1968).

7. Pizon, V., Cuny, G. & Bernadi, G. *Eur. J. Biochem.* **140**, 25–30 (1984).
8. Smith, C. L., Econome, J. G., Schutt, A., Kico, S. & Cantor, C. R. *Science* **236**, 1448–1453 (1987).
9. Hinegardner, R. & Rosen, D. E. *Am. Nat.* **106**, 621–644 (1972).
10. Moore, S. S. et al. *Genomics* **10**, 654–660 (1991).
11. Beckmann, J. S. & Weber, J. L. *Genomics* **12**, 627–631 (1992).
12. Stallings, R. L. et al. *Genomics* **10**, 807–815 (1991).
13. Lanza, F., Kieffer, N., Phillips, D. R. & Fitzgerald, L. A. *J. biol. Chem.* **265**, 18098–18103 (1990).
14. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
15. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. *J. molec. Biol.* **215**, 403–410 (1990).
16. Parsons, J. D., Brenner, S. & Bishop, M. J. *CABIOS* **8**, 461–466 (1992).
17. Devereux, J. R., Haeberli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387–395 (1984).
18. Karn, J., Matthes, H. W. D., Gait, M. J. & Brenner, S. *Gene* **32**, 217–224 (1984).
19. Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991–1994 (1984).

ACKNOWLEDGEMENTS. R.S., A.M. and S.A. are in receipt of MRC Training Fellowships. The work was supported by the Louis Jeantet Foundation (including a fellowship for G.E.) and by funds from E. I. du Pont de Nemours Ltd. We thank Dr Y. Masuda of Kagoshima University, Japan, for help in obtaining *Fugu* material and K. Hawker and G. Lye for help with sequencing.

Novel heterotrimeric kinesin-related protein purified from sea urchin eggs

D. G. Cole, S. W. Chinn, K. P. Wedaman, K. Hall, T. Vuong & J. M. Scholey*

Section of Molecular and Cellular Biology, Division of Biological Sciences, University of California, Davis, California 95616, USA

* To whom correspondence should be addressed

KINESIN heavy chain and kinesin-related polypeptides (KRPs) comprise a family of motor proteins with diverse intracellular transport functions^{1–7}. Using pan-kinesin peptide antibodies that react with these proteins^{8,9}, we have previously purified from sea urchin eggs a trimeric microtubule-binding and bundling protein, KRP_(85/95) (ref. 8) comprising subunits of *M_r* 115,000 (115K), 95K and 85K. We report here that kinesin-related genes encode the 85K and 95K subunits, and that the protein can be immunoprecipitated from cytosol as a trimeric complex using an 85K monoclonal antibody. We also find that purified KRP_(85/95) directs movements towards the 'plus' ends of microtubules. To our knowledge, this protein is the first kinesin-related motor to be purified from its natural host cell in a native multimeric state.

We prepared KRP_(85/95) monoclonal immunoglobulin Gs against the 95K (antibody K2.1) and 85K subunits (antibodies

FIG. 1 Co-immunoprecipitation of the three subunits of KRP_(85/95) using monoclonal antibody resin. **a**, Coomassie-blue-stained SDS-polyacrylamide gels of purified sea urchin egg kinesin (lane 1) and KRP_(85/95) (lane 2; ref. 8). Subunit *M_s* (in thousands) are indicated. **b**, Precipitation of KRP_(85/95) from cytosol using the anti-85K monoclonal antibody (mAb) K2.2. The gel (lanes 1–3) and corresponding immunoblot probed with a mouse KRP_(85/95) antiserum (lanes 1'–3') show egg cytosol (lanes 1, 1') and the immunoprecipitates obtained using anti-85K IgG (lanes 3, 3') and a control anti- β -tubulin IgG (lanes 2, 2'). The heavy 50K bands (lanes 2, 3, 2', 3') are the γ -chains of the two mAbs that reacted with anti-mouse secondary antibody (lanes 2', 3'). The 115K, 95K and 85K subunits of KRP_(85/95) are indicated. Immunoprecipitation procedure is described in ref. 20. **c**, Immunoblots of egg cytosol (left; see lane 1 in **b**), ATP-eluted microtubule-associated proteins (MAPs) from egg AMP-PNP MTs (centre; ref. 8), and purified KRP_(85/95) (right; see lane 2 in **a**) probed with mouse antiserum against the KRP (lanes 1, 6, 11), with mAb K2.1 to the 95K subunit (lanes 2, 7, 12), and with three 85K mAbs, K2.2 (lanes 3, 8, 13) K2.3 (lanes 4, 9, 14) and K2.4 (lanes 5, 10, 15).

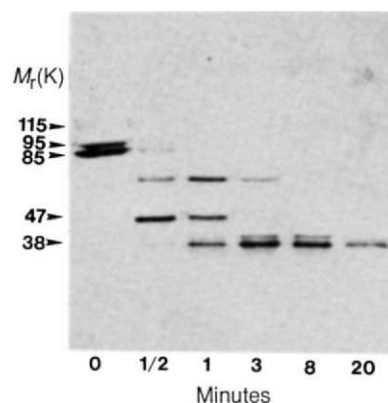
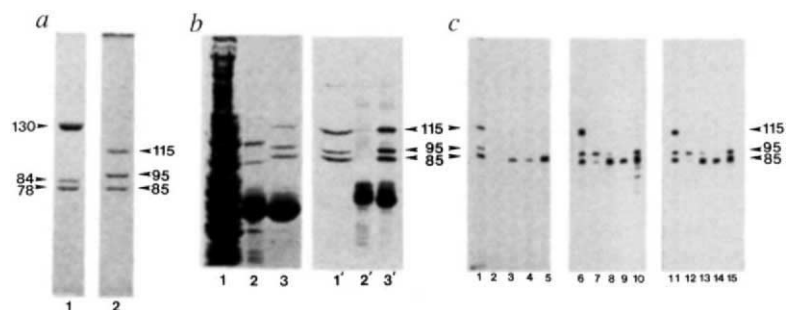


FIG. 2 Production of motor domain subfragments of KRP_(85/95) for N-terminal microsequencing. Purified KRP_(85/95) was digested with chymotrypsin to yield the motor domain fragments (as described for KHC; refs 10, 11) and the digests were probed with pankinesin peptide antibody⁸ to detect 47K and 38K 'motor' fragments of the 85K and 95K subunits as indicated (the 115K 'accessory' subunit does not react; ref. 8). Numbers under each lane show length of digestion time before quenching. Duplicate samples were transferred to Problot membranes, Coomassie-blue-stained, and the 38K and 47K motor fragments excised and partially sequenced. The 47K fragment behaves as a precursor of the 38K peptide. The N-terminal ten residues of both fragments were found to be PGGXSXGNDNV, where X is an unknown amino acid. The yield of each residue was 2–10 pmol from an unquantitated amount of starting material.

K2.2, K2.3, K2.4) from mice immunized with highly purified KRP_(85/95) (Fig. 1a). In support of the hypothesis that the 115K, 95K and 85K polypeptides of KRP_(85/95) are bound together as a complex in cells, we observed that K2.2 was able to immunoprecipitate the three polypeptides directly from the cytosol (Fig. 1b).

On immunoblots the four monoclonals reacted specifically with subunits of KRP_(85/95) but not with other polypeptides present in egg cytosol (Fig. 1c), including those that co-precipitated with microtubules from cytosol treated with AMP-PNP, a non-hydrolysable ATP analogue, and were eluted from these microtubules by ATP (Fig. 1c). The lack of crossreaction with kinesin heavy chain and other KRPs present in egg AMP-PNP-microtubules⁸ suggests that these antibodies can discriminate KRP_(85/95) subunits from other relatives of kinesin. We observed

TABLE 1 Microtubule motion driven by kinesin and KRP_(85/95)

	Kinesin	KRP _(85/95)
Control	0.66+/-0.03 $\mu\text{m s}^{-1}$ (n=10)	0.41+/-0.04 $\mu\text{m s}^{-1}$ (n=10)
+ SUK4	0.01+/-0.02 $\mu\text{m s}^{-1}$ (n=10)	0.42+/-0.05 $\mu\text{m s}^{-1}$ (n=10)
+ K2.2	0.65+/-0.04 $\mu\text{m s}^{-1}$ (n=10)	No MTs attached
+ K2.3	0.66+/-0.05 $\mu\text{m s}^{-1}$ (n=10)	No MTs attached

Motility assays were done with purified kinesin or KRP_(85/95) (Fig. 1a) in the presence or absence of purified IgGs as described in ref. 10 for kinesin. The polarity of microtubule (MT) motion driven by KRP_(85/95) was determined with fluorescent 'marked' MTs prepared using rhodamine-tubulin, N-ethylmaleimide-modified tubulin, GMP-CPP (α , β -methylene guanosine-5'-triphosphate) and oxygen scavengers (gift from R. D. Vale) according to ref. 19. The polarity was assayed on two preparations of purified KRP_(85/95); in both cases 100% of MTs (n>100) moved with their minus ends leading (both in the presence and absence of SUK4), demonstrating that KRP_(85/95) is a plus-end-directed MT motor.

to determine the sequence of their ten N-terminal amino-acid residues as PGGSGXGNDNV. This sequence was used to design corresponding oligonucleotide primers which were used in the polymerase chain reaction to amplify the corresponding motor domain complementary DNA from a pool enriched in cDNAs that encode sea urchin egg KRPs¹³. The product was used to clone sea urchin egg cDNAs spanning the entire open reading frame (ORF) encoding one kinesin-related subunit of KRP_(85/95) and part of a very similar ORF encoding the other (Fig. 3). By comparing the sequences deduced from these ORFs with sequences of peptides derived from the three electrophoretically separated KRP_(85/95) subunits, we demonstrated that the complete and partial ORFs, respectively, encode the full-length 85K subunit and a portion of the 95K subunit.

The complete deduced sequence of the 85K subunit was analysed in detail: it predicts a 699-amino-acid polypeptide with M_r 78,763. There is an N-terminal motor domain (residues 1–340; predicted M_r 37,398; pI 9.7) which is 46% identical with the sea urchin kinesin heavy chain motor domain (Fig. 3c), and contains the consensus residues present in the motor domains of all KRPs so far described^{5,13}. As with kinesin heavy chain, the motor domain is separated from a small, basic, globular C-terminal domain (residues 620–699; predicted pI 10.8) by a central α -helical segment (residues 341–619) that is predicted to form a coiled coil (Fig. 3d) and may be the site of dimerization with the 95K subunit of KRP_(85/95). When the 85K ORF was compared with other predicted KRP sequences, we found the highest level of identity with *Drosophila* KLP4 (ref. 13) and mouse KIF3 (ref. 14); overall, there was 33% identity and 56% similarity with kinesin heavy chain, 56% identity and 70% similarity with KLP4, and 71% identity and 86% similarity with KIF3. Studies of the polypeptides encoded by the KLP4 and KIF3 genes have not been reported, but our data suggest that they are likely to be components of trimeric complexes in their natural host cells.

To confirm that KRP_(85/95) is indeed a microtubule-based motor protein, it was important to demonstrate that the purified protein can drive microtubule movement. In a motility assay^{2,15}, purified KRP_(85/95) showed plus-end-directed motor activity like kinesin itself (Table 1). To eliminate the possibility that this activity might be due to a low level of contamination with kinesin, we used function-blocking antibodies (Table 1). Both kinesin and KRP_(85/95) induced microtubule motility at rates of 0.6–0.7 and 0.4 $\mu\text{m s}^{-1}$, respectively, in the absence of antibody. The anti-kinesin SUK4 antibody¹⁰ eliminated kinesin-driven motility but did not effect motility driven by KRP_(85/95). Conversely, the KRP_(85/95)-specific antibodies K2.2 and K2.3 had no effect on kinesin-driven motility, but completely blocked the activity of KRP_(85/95), inhibiting microtubules from attaching to the motor-coated surface (as observed previously with a polyclonal kinesin antibody¹⁶). Therefore, KRP_(85/95) has *bona fide* motor activity.

KRP_(85/95) is, to our knowledge, the first relative of kinesin to be directly purified from its natural host cell. Much remains to be learned about the structure, enzymatic properties and intracellular functions of this new heterotrimeric motor protein. Of particular interest is the role of the 115K subunit, which is apparently not a kinesin-related polypeptide but may be analogous to kinesin light chains^{17,18}; this 'accessory' polypeptide may function as an adaptor for cargo attachment or as regulatory subunit. □

Received 26 August; accepted 23 September 1993.

1. Yang, J. T., Laymon, R. A. & Goldstein, L. S. B. *Cell* **56**, 879–889 (1989).
2. Vale, R. D., Reese, T. S. & Sheetz, M. P. *Cell* **42**, 39–50 (1985).
3. Brady, S. T. *Nature* **317**, 73–75 (1985).
4. Scholey, J. M., Porter, M. E., Grissom, P. M. & McIntosh, J. R. *Nature* **318**, 483–486 (1985).
5. Goldstein, L. S. B. *Trends Cell Biol.* **1**, 93–98 (1991).
6. Vale, R. D. *Trends biochem. Sci.* **17**, 300–304 (1992).
7. Skoufias, D. A. & Scholey, J. M. *Curr. Opin. Cell Biol.* **5**, 95–104 (1993).
8. Cole, D. G. et al. *J. Cell Sci.* **101**, 291–301 (1992).
9. Sawin, K. E., Mitchison, T. J. & Wordeman, L. G. *J. Cell Sci.* **101**, 303–313 (1992).
10. Ingold, A. L., Cohn, S. A. & Scholey, J. M. *J. Cell Biol.* **107**, 2657–2667 (1988).
11. Kuznetsov, S. A. et al. *J. biol. Chem.* **264**, 589–595 (1989).
12. Scholey, J. M., Heuser, J., Yang, J. T. & Goldstein, L. S. B. *Nature* **338**, 355–357 (1989).
13. Stewart, R. J., Pesavento, P. A., Woerpel, D. N. & Goldstein, L. S. B. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8470–8474 (1991).
14. Aizawa, A. et al. *J. Cell Biol.* **119**, 1287–1296 (1992).
15. Cohn, S. A., Ingold, A. L. & Scholey, J. M. *J. biol. Chem.* **264**, 4290–4297 (1989).
16. Rodionov, V. I., Gyoeva, F. K. & Gelfand, V. I. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4956–4960 (1991).
17. Cyr, J. L., Pfister, K. K., Bloom, G. S., Slaughter, C. A. & Brady, S. T. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10114–10118 (1991).
18. Wedaman, K. P., Knight, A. E., Kendrick-Jones, J. & Scholey, J. M. *J. molec. Biol.* **231**, 155–158 (1993).
19. Hyman, A. A. & Howard, J. *Meth. Cell Biol.* **39**, (in the press).
20. Johnson, C. S., Buster, D. & Scholey, J. M. *Cell Motil. Cytoskel.* **16**, 204–213 (1990).
21. Wright, B. D. et al. *J. Cell Biol.* **113**, 817–833 (1991).
22. Patterson, S. D., Hess, D., Yungwirth, T. & Aebersold, R. *Analyt. Biochem.* **202**, 193–203 (1992).

ACKNOWLEDGEMENTS. We thank members of our laboratory, R. Baskin, A. El-Badry, R. Scibien-ski, C. McNeil, the Vale laboratory and G. Munske for their help. L. S. B. Goldstein for the KLP4 sequence, G. Munske and Univ. Calif. at Davis Protein Structure laboratory for amino-acid sequencing, and S. Brady for discussion. This work was supported by grants from the NIH and ACS to J.M.S., a Presidential Undergraduate Fellowship to S.W.C., and a NIH postdoctoral fellowship to D.G.C. The sequences of the 85K and 95K subunits are available from Genbank under accession numbers L16993 and U00996.

Calmodulin-dependent protein kinase II mediates inactivation of MPF and CSF upon fertilization of *Xenopus* eggs

Thierry Lorca*, Francisco H. Cruzalegui†, Didier Fesquet*, Jean-Claude Cavadore*, Jean Méry*, Anthony Means‡ & Marcel Dorée*

* Centre National de la Recherche Scientifique and Institut National de la Santé et de la Recherche Médicale, BP 5051, 34033 Montpellier Cedex, France

† Department of Cell Biology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

‡ Department of Pharmacology, Duke University Medical Center, Durham, North Carolina 27710, USA

IN vertebrates, unfertilized eggs are arrested at second meiotic metaphase by a cytostatic factor (CSF)¹, an essential component of which is the product of the *c-mos* proto-oncogene². CSF prevents ubiquitin-dependent degradation of mitotic cyclins and thus inactivation of the M phase-promoting factor (MPF)^{3,4}. Fertilization or parthenogenetic activation triggers a transient increase in the cytoplasmic free Ca^{2+} (reviewed in refs 5 and 6), inactivates both CSF and MPF, and releases eggs from meiotic metaphase arrest^{7–10}. A calmodulin-dependent process is required for cyclin

degradation to occur in cell-free extracts prepared from metaphase II-arrested eggs (CSF extracts) when the free Ca^{2+} concentration is transiently raised in the physiological micromolar range¹⁰. Here we show that when a constitutively active mutant of calmodulin-dependent protein kinase II (CaM K_{II}) is added to a CSF extract, cyclin degradation and Cdc2 kinase inactivation occur even in the absence of Ca^{2+} , and the extract loses its ability to cause metaphase arrest when transferred into embryos. Furthermore, specific inhibitors of CaM K_{II} prevent cyclin degradation after calcium addition. Finally, the direct microinjection of constitutively active CaM K_{II} into unfertilized eggs inactivates Cdc2 kinase and CSF, even in the absence of a Ca^{2+} transient. The target for Ca^{2+} -calmodulin is thus CaM K_{II} .

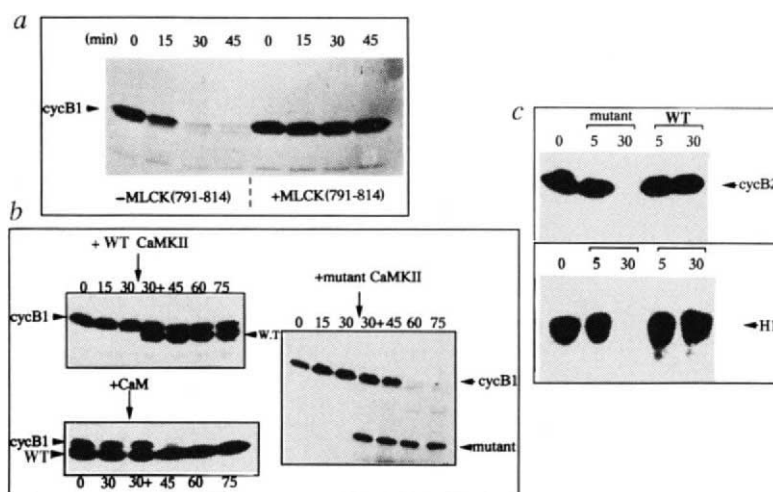
We previously reported that a calmodulin-binding peptide from myosin light chain kinase^{11,12}, MLCK(791–814), prevents exogenous Ca^{2+} from triggering cyclin degradation in CSF extracts (ref. 10, and Fig. 1a). This suggests that Ca^{2+} first activates a calmodulin-dependent enzyme. As ablation of limited domains of calmodulin-dependent enzymes often renders them independent of Ca^{2+} for catalytic activity^{13–17}, we tested the *in vitro* action of a constitutively active mutant of CaM K_{II} (fragment 1–290 of rat brain CaM kinase II α ; ref. 17) in this system. CaM K_{II} (1–290) readily released the cyclin degradation machinery from its MLCK(791–814)-inhibited state, even in the absence of Ca^{2+} (Fig. 1b and c). In contrast, the holoenzyme was ineffective in such conditions, even after Ca^{2+} addition. To show

that this was due to the presence of MLCK(791–814), we added excess calmodulin to CSF extracts to counteract the effect of MLCK(791–814): cyclin B was rapidly degraded (Fig. 1b). Microinjection into metaphase II-arrested oocytes of the constitutively active CaM K_{II} (together with 5 mM EGTA to prevent parthenogenetic activation by pricking) also inactivated Cdc2 kinase (MPF) and triggered *c-mos* degradation, whereas the holoenzyme had no effect (Fig. 2a and b). Moreover, the ability of egg extracts to arrest the cell cycle at metaphase when microinjected into a blastomere of a two-cell embryo (CSF activity)¹ also disappeared upon addition of the truncated CaM K_{II} , but not of its non-truncated counterpart, and embryos receiving cytoplasm taken from unfertilized eggs injected with constitutively active CaM K_{II} failed to arrest (Fig. 2c).

Although these results suggested that CaM K_{II} had caused CSF inactivation, an alternative interpretation could have been that coinjecting unregulated CaM K_{II} with CSF into a blastomere overcomes its cytostatic effect. In order to suppress CaM K_{II} activity in extracts before microinjection, we used CaM K_{II} (281–302), a conserved peptide¹⁸ corresponding to the auto-inhibitory domain of CaM K_{II} that specifically inhibits this enzyme (as well as its constitutively activated deletion mutants)^{19–22}. As shown in Fig. 3a and b, CaM K_{II} (281–302) readily suppressed cyclin degradation when it was added before Ca^{2+} or the constitutively active CaM K_{II} mutant, respectively. When the inhibitory peptide was added after cyclin degradation

FIG. 1 Ca^{2+} -independent activation of cyclin degradation in CSF extracts by a constitutively active CaM K_{II} deletion mutant. a, Autoradiogram monitoring degradation of [³⁵S]methionine-labelled human cyclin B1 as a function of time elapsed from Ca^{2+} addition in a CSF extract (30 μl) with (right) or without (left) the MLCK(791–814) peptide (0.1 mg ml^{-1}). Radiolabelled cyclin (1 μl) was added simultaneously with the inhibitory peptide, 5 min before Ca^{2+} addition. b, Degradation of cyclin B1 upon addition of a truncated mutant of CaM K_{II} ending at Leu 290 (right) or the full-length enzyme (left; WT, wild type). In both cases the MLCK(791–814) peptide was added at zero time with cyclin B1. Upper left and right panels, CaM K_{II} , translated in reticulocyte lysate, was added (4 μl per 30 μl extract) at 30 min (arrow), in the presence (WT holoenzyme) or absence (CaM K_{II} (1–290) mutant) of Ca^{2+} ; lower left panel, Ca^{2+} was added at zero time in a CSF extract containing WT CaM K_{II} , cyclin B1 and the MLCK(791–814) peptide. Thirty minutes later (arrow), calmodulin (CaM) was added in excess (17 μM) together with 50 μM Ca^{2+} to counteract the inhibitory effect of the MLCK(791–814) peptide. c, Effect of adding either the CaM K_{II} (1–290) mutant or the full-length enzyme on degradation of endogenous cyclin B2 or inactivation of H1 histone kinase activity in CSF extracts containing the MLCK(791–814) peptide. Mutant or wild-type enzymes were added (4 μl) in 30 μl CSF extract (no Ca^{2+} was added). Cdc2-containing complexes were isolated at the times indicated from aliquots by affinity chromatography on p13^{suc1} beads. They were analysed by immunoblotting for detection of *Xenopus* cyclin B2 (top) or used to phosphorylate H1 histones in the presence of [γ -³²P]ATP (bottom).

METHODS. Plasmids encoding rat brain CaM kinase II α -holoenzyme (pET3a) and the mutant ending at Leu 290 (pT7–7) were constructed as described¹⁷. pET3a and pT7–7 plasmids were linearized using *Bam*HI and *Pst*II, respectively, then transcribed with T7 RNA polymerase. For *in vitro* translation reactions, 1 μg of each transcript was used in 15 μl rabbit reticulocyte lysate containing 10 μCi [³⁵S]methionine. Translation was allowed to proceed for 50 min at 30 °C. The calcium-independent activity of the translation mixture containing the holoenzyme was lower than 0.3 $\text{pmol min}^{-1} \mu\text{l}^{-1}$ of lysate (using syntide 2 as a substrate; see Fig. 4), and it increased to about 5 $\text{pmol min}^{-1} \mu\text{l}^{-1}$ in the presence of Ca^{2+} . In comparison, the calcium-independent activity of the *in vitro*-translated mutant was about 3.5 $\text{pmol min}^{-1} \mu\text{l}^{-1}$ of lysate. The full-length human cyclin B1, subcloned into the *Bam*HI site of pGEM4Z (Promega) was a gift from J. Pines. It was transcribed by SP6 polymerase and translated in reticulocyte lysate containing



[³⁵S]methionine as described²⁷. Solid-phase synthesis of the MLCK(791–814) peptide (MKKYMARRKWQKTGHAVRAIGRLS) as well as other peptides used in this study was performed on a Milligen 9050 peptide synthesizer using fluorenylmethoxycarbonyl (Fmoc) groups as a temporary amino protection, and purified by HPLC on an Aquapore RP300 column. Calmodulin was purified to apparent homogeneity from ram testis and CSF extracts were prepared from unfertilized *Xenopus* eggs as described¹⁰. They contained 7 mM EGTA; thus, addition of 0.6 mM CaCl_2 (in a) increased free Ca^{2+} only to the micromolar range, as checked with a specific Ca^{2+} microelectrode (results not shown). Degradation of human cyclin B1 was assayed by running aliquots on 10% polyacrylamide gels. Gels were treated with an enhancer for low-energy radiation (DuPont), dried and kept at -70°C for fluorography. To isolate Cdc2-containing complexes (c) 15- μl aliquots of CSF extract were diluted in 500 μl RIPA buffer¹⁰ containing 25 μl p13^{suc1}-Sepharose beads (3.5 μg yeast protein bound per μl). Pellets were washed and assayed for H1 histone kinase activity as before¹⁰. After kinase assay, eluted materials and H1 histones were run together on the same polyacrylamide gel, transferred on a nitrocellulose sheet, and analysed by ECL western blotting (Amersham, UK) with a polyclonal antiserum directed against *Xenopus* cyclin B2 (a gift from M. Glotzer) and by autoradiography to detect H1 histone phosphorylation.

FIG. 2 Unregulated CaM $K_{II}(1-290)$ inactivates Cdc2 kinase and CSF in the absence of Ca^{2+} . **a**, Reticulocyte lysates containing 5 mM EGTA and either the CaM $K_{II}(1-290)$ fragment or the CaM K_{II} holoenzyme were microinjected (50 nl) into *in vitro*-matured oocytes arrested at meiotic metaphase II (after overnight treatment with 2 μ M progesterone). Thirty minutes after microinjection, 10 recipient oocytes were crushed in 50 μ l of a buffer containing 50 mM Tris, pH 7.5, 150 mM NaCl and 1 mM EGTA, centrifuged for 2 min at 10,000g and the supernatant immunoprecipitated with 5 μ l of an antiserum directed against the 16-amino-acid C-terminal sequence of *Xenopus* Cdc2 after dilution with 1 ml RIPA buffer¹⁰. H1 kinase activity was measured as described²⁸ in immune complexes isolated on protein A-Sepharose. Cdc2 kinase activity is represented by bars as pmol phosphate transferred to H1 histone per min per oocyte. **b**, Oocytes were matured in the presence of 1 mCi ml⁻¹ [³⁵S]methionine, and p39^{mos} immunoprecipitated using an antiserum directed against bacterially produced p39^{mos}, then analysed by gel electrophoresis and fluorography¹⁰. Mature oocytes were injected with wild-type CaM K_{II} (left); or with CaM $K_{II}(1-290)$ (right). **c**, *In vitro*-translated CaM $K_{II}(1-290)$ fragment or CaM K_{II} holoenzyme were added (4 μ l) to CSF extracts (30 μ l; **A**) or microinjected (30 nl, together with 5 mM EGTA) into unfertilized eggs (**B**). After 30 min, extracts (**A**) or cytoplasm taken from injected eggs (**B**) were transferred into one of the blastomeres of two-cell embryos (30 nl per blastomere). Microinjected blastomeres that arrested (either without dividing or after dividing only once) were scored when the non-injected blastomere of the two-cell embryo had generated more than 16 cells (about 2 h after fertilization). CSF activity is represented by bars as per cent arrest of recipient embryos. The numbers of recipient embryos and those undergoing arrest are also indicated.

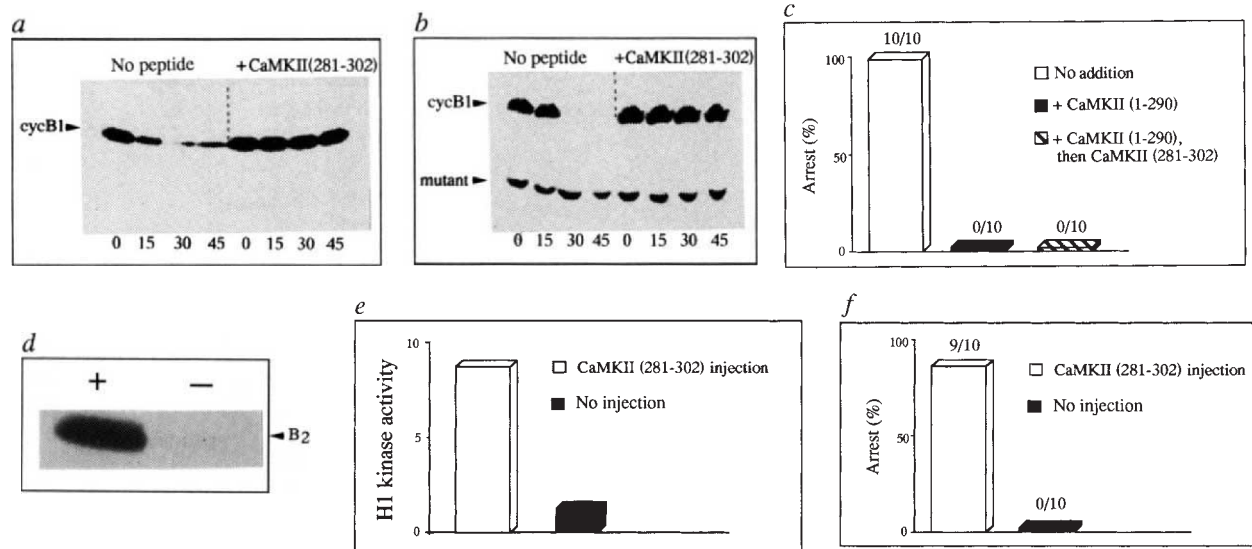
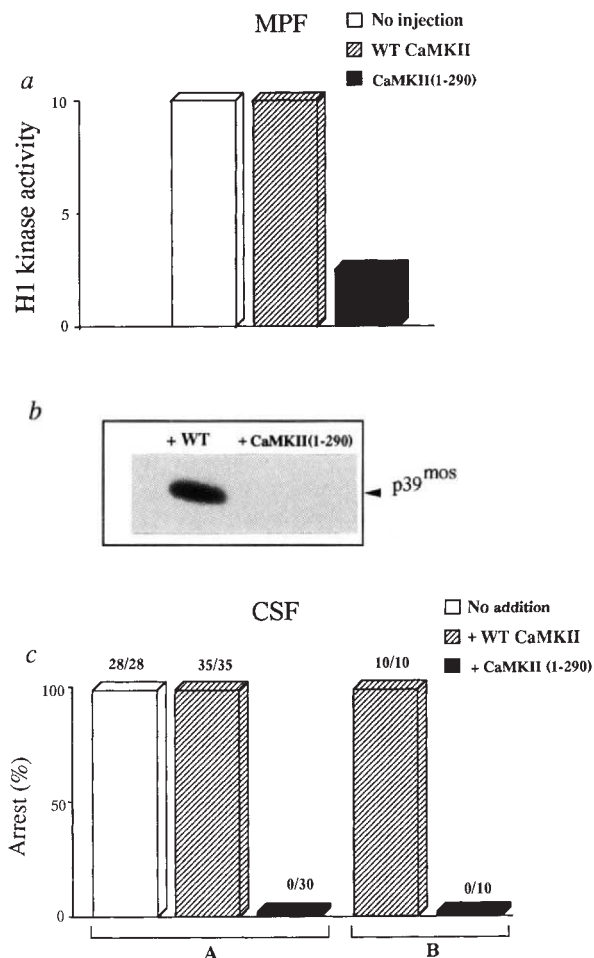
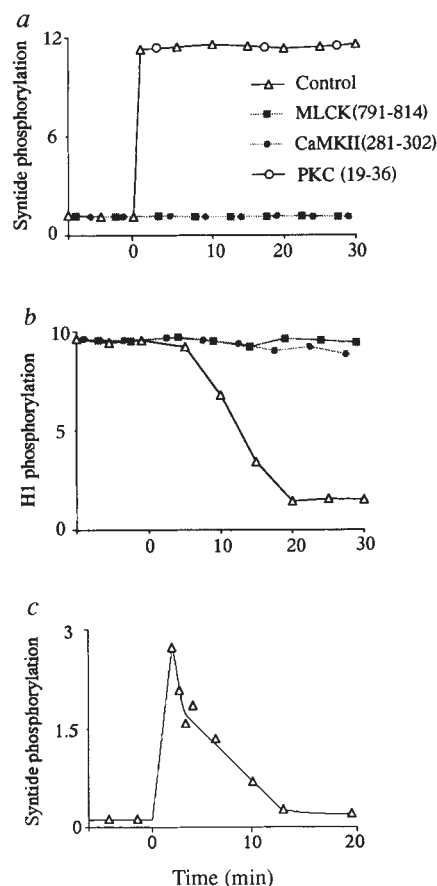


FIG. 3 CaM K_{II} mediates cyclin degradation and inactivation of both Cdc2 kinase and CSF. **a**, Degradation of [³⁵S]methionine-labelled human cyclin B1 in CSF extracts with or without (control) 0.1 mg ml⁻¹ CaM $K_{II}(281-302)$ peptide (MHRQETVDCLKKFNARRKLKGA), as a function of time (min) elapsed from Ca^{2+} addition (see Fig. 1 legend). **b**, As **a**, but with no Ca^{2+} added; cyclin degradation was triggered by the addition of *in vitro*-translated CaM $K_{II}(1-290)$ as described in Fig. 1 legend. **c**, *In vitro*-translated CaM $K_{II}(1-290)$ fragment or CaM K_{II} holoenzyme were added (4 μ l) to CSF extracts (30 μ l). Thirty min later (after cyclin degradation), the CaM $K_{II}(281-302)$ peptide was added (0.1 mg ml⁻¹), then extracts were transferred into one of the blastomeres of two-cell embryos (30 nl per blastomere) to assay CSF activity (Fig. 2c

legend) as per cent arrest of the recipient embryos. **d-f**, Unfertilized eggs were microinjected with CaM $K_{II}(281-302)$ inhibitory peptide (50 nl, 10 mg ml⁻¹) using the method of Wang²⁹ to avoid activation by pricking; they then received a brief electric shock to activate them, and 30 min later were analysed for cyclin degradation (**d**), Cdc2 kinase inactivation (**e**) or CSF inactivation (**f**). In **d** and **e**, homogenates were prepared from 10 injected or 10 control non-injected eggs and analysed as described in Fig. 1 legend for degradation of cyclin B2 (plus sign, injected eggs; minus sign, control eggs) and as described in Fig. 2 legend for determination of Cdc2 kinase activity, respectively. In **f**, cytoplasm (50 nl) was taken from injected or non-injected eggs and assayed for CSF activity.

FIG. 4 Time course of endogenous CaM K_{II} activation monitored using syntide 2, a specific target peptide. *a*, The Ca^{2+} -dependent activity that phosphorylates syntide 2 was monitored (pmol phosphate transferred per min per μ l extract) as a function of time elapsed from Ca^{2+} addition in CSF extracts with or without (control) 0.1 mg ml^{-1} potential peptide inhibitors (MLCK(791-814); CaM K_{II} (281-302); PKC(19-36)). *b*, H1 histone kinase activity was monitored (pmol of phosphate transferred per min per μ l extract) under the same conditions as in *a* (same symbols). *c*, Calcium-independent phosphorylation of syntide 2 in extracts prepared at the indicated time after parthenogenetic activation from electrically stimulated eggs.

METHODS. For *in vitro* experiments (*a* and *b*), 2- μ l aliquots were taken at the indicated times and transferred either into 25 μ l of a mixture containing 70 mM HEPES, pH 7.4, 4 mM $MgCl_2$, 0.2 mM $[\gamma\text{-}^{32}P]ATP$ (125 c.p.m. per picomol), and 1 mg ml^{-1} of the syntide 2 peptide (*a*) or into 40 μ l of a mixture containing 40 mM HEPES, pH 7.4, 10 mM $MgCl_2$, 0.2 mM $[\gamma\text{-}^{32}P]ATP$ (125 c.p.m. per pmol) and 1 mg ml^{-1} of H1 histone (*b*). After 10 min incubation at 25 $^{\circ}C$, samples were spotted onto phosphocellulose papers, then extensively washed in tap water. Incorporated radioactivity was determined by liquid scintillation counting. Phosphoaminoacid analysis confirmed that Ca^{2+} -dependent phosphorylation of syntide 2 occurred exclusively on its serine residue (not shown). To monitor *in vivo* activation of CaM K_{II} (*c*), dejellied eggs were crushed at the indicated times after parthenogenetic activation by electric shock in 25 μ l of a mixture containing 70 mM HEPES, pH 7.4, 4 mM $MgCl_2$, 7 mM EGTA, 0.2 mM $[\gamma\text{-}^{32}P]ATP$ (125 c.p.m. per pmol) and either the syntide 2 peptide (1 mg ml^{-1}) or not (background). After 3 min incubation at 25 $^{\circ}C$, samples were processed as described to determine calcium-independent phosphorylation of syntide 2.



to inactivate CaM K_{II} before CSF activity of extracts was assayed, microinjected blastomeres still failed to arrest (Fig. 3*c*). Finally, unfertilized eggs microinjected with the CaM K_{II} (281-302) inhibitory peptide failed to undergo cyclin degradation (Fig. 3*d*) and inactivation of either Cdc2 kinase (Fig. 3*e*) or CSF (Fig. 3*f*) upon electrical stimulation, a treatment that causes parthenogenetic activation in non-injected eggs. Taken together, these results indicate that CaM K_{II} mediates the Ca^{2+} -dependent inactivation of MPF and CSF that accompanies fertilization or egg activation.

To demonstrate the activation of CaM K_{II} by calcium in CSF extracts directly, we used syntide 2, a peptide with a sequence (PLARTLSVAGLPGKK) similar to that of phosphorylation site 2 of glycogen synthase²³, which contains the consensus motif (RXX^S_{/I}) for CaM K_{II} (refs 24, 25). When Ca^{2+} was added to CSF extracts, an activity that phosphorylates this peptide on serine increased dramatically (Fig. 4*a*), and earlier than the drop of H1 kinase activity associated with cyclin degradation (Fig. 4*b*). The synthetic peptide CaM K_{II} (281-302) suppressed the

burst of syntide 2 phosphorylation following Ca^{2+} addition, whereas protein kinase C(19-36), corresponding to the auto-inhibitory domain of another Ca^{2+} -dependent kinase²⁶, had no effect (Fig. 4*a*). These results identify the kinase responsible for syntide 2 phosphorylation in CSF extracts as CaM K_{II} . An early and transient activation of CaM K_{II} was also observed in intact eggs after parthenogenetic activation (Fig. 4*c*).

Our results show that activation of CaM K_{II} is both necessary and sufficient for CSF to be inactivated and for the cyclin degradation machinery to be released from its inhibited state. But cyclin degradation occurs several minutes before CSF activity drops and before any changes in level or size of the *Xenopus c-mos* product can be detected following activation of *Xenopus* eggs by pricking or using a calcium ionophore⁸. It is therefore possible that Ca^{2+} , as a result of activation of CaM K_{II} , first overcomes inhibition of cyclin degradation by CSF, and only then causes CSF inactivation. The molecular mechanisms by which CaM K_{II} governs MPF and CSF inactivation, and the exact relationship between the two events, are under investigation. □

Received 26 April; accepted 12 August 1993.

- Masui, Y. & Markert, C. L. *J. exp. Zool.* **177**, 129-146 (1971).
- Sagata, N., Watanabe, N., Vande Woude, G. F. & Ikawa, Y. *Nature* **342**, 512-518 (1989).
- Glötzer, M., Murray, A. W. & Kirschner, M. W. *Nature* **349**, 132-138 (1991).
- Murray, A. W., Solomon, M. J. & Kirschner, M. W. *Nature* **339**, 280-286 (1989).
- Jaffe, L. *Dev. Biol.* **99**, 265-276 (1983).
- Berridge, M. J. *Nature* **361**, 315-325 (1993).
- Meyerohof, P. G. & Masui, Y. *Dev. Biol.* **61**, 214-229 (1977).
- Watanabe, N., Hunt, T., Ikawa, Y. & Sagata, N. *Nature* **352**, 247-248 (1991).
- Weber, M., Kubiak, J. Z., Arlinghaus, R. B., Pines, J. & Maro, B. *Dev. Biol.* **148**, 393-397 (1991).
- Lorca, T. et al. *EMBO J.* **10**, 2087-2093 (1991).
- Blumenthal, D. K. et al. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3187-3191 (1985).
- Olson, N. J. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2284-2288 (1990).
- Cheung, W. Y. *J. Biol. Chem.* **246**, 2859-2869 (1971).
- Depaoli-Roach, A. A., Gibbs, J. B. & Roach, P. *FEBS Lett.* **105**, 321-324 (1979).

- Niggli, U., Adunyah, E. S. & Carafoli, E. *J. Biol. Chem.* **256**, 8588-8592 (1981).
- Manalan, A. S. & Klee, C. B. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4291-4295 (1983).
- Cruzalegui, F. H. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 12127-12131 (1992).
- Pausch, M. H., Kaim, D., Kunisawa, R., Admon, A. & Thörner, J. *EMBO J.* **10**, 1511-1522 (1991).
- Söderling, T. R. *J. Biol. Chem.* **265**, 1823-1826 (1990).
- Payne, M. E. et al. *J. Biol. Chem.* **263**, 7190-7195 (1988).
- Malinow, R., Schulman, H. & Tsien, R. W. *Science* **245**, 862-866 (1989).
- Schulman, H. & Lou, L. L. *Trends biochem. Sci.* **14**, 62-66 (1989).
- Colbran, R. J., Fong, Y. L., Schworer, C. M. & Söderling, T. R. *J. Biol. Chem.* **263**, 18145-18151 (1988).
- Pearson, R. B. & Kemp, B. E. *Meth. Enzym.* **200**, 62-81 (1991).
- Woodgett, J. R., Davison, M. T. & Cohen, P. *Eur. J. Biochem.* **136**, 481-488 (1983).
- Smith, M. K., Colbran, R. J. & Söderling, T. R. *J. Biol. Chem.* **265**, 1837-1840 (1990).
- Lorca, T. et al. *J. Cell Sci.* **102**, 55-62 (1992).
- Labbé, J. C., Cavadore, J. C. & Dorée, M. *Meth. Enzym.* **200**, 291-301 (1991).
- Wangh, L. J. *J. Cell Sci.* **93**, 1-8 (1989).

Bcl-2 associates with the ras-related protein R-ras p23

Maria Jose Fernandez-Sarabia
& James R. Bischoff

ONYX Pharmaceuticals, 3031 Research Drive, Building A,
Richmond, California 94806, USA

APOPTOSIS is an important but poorly understood mechanism of cell regulation. Growth factor deprivation can trigger apoptosis in a variety of cells¹⁻³, suggesting the existence of a signal transduction pathway responding to external signals and leading to apoptosis. Overexpression of the proto-oncogene *bcl-2* can override these signals and block apoptosis⁴⁻¹⁴, indicating that the *bcl-2* protein (Bcl-2) is an important component of the apoptotic response. The identification of Bcl-2-binding proteins might help explain how Bcl-2 acts to regulate apoptosis. Here we use the yeast two-hybrid system¹⁵ to show that the human *ras*-related protein R-ras p23 (refs 16-18) binds to Bcl-2. This association is also detected in immunoprecipitates from human cell extracts. The association requires full-length Bcl-2 but the C-terminal 60 amino acids of R-ras p23 are sufficient for the interaction. These results provide evidence of a putative component of a signal transduction pathway involved in the regulation of apoptosis.

We used the yeast two-hybrid system¹⁵ to isolate complementary DNAs encoding proteins that bind to Bcl-2. Three independently isolated clones, clones 1-3, were found to encode C-terminal fragments of the human *ras*-related protein R-ras (Fig. 1a). Human R-ras encodes a 218-amino-acid polypeptide (p23) which is 55% identical to H-ras p21 (ref. 16). Although the structure of R-ras p23 is similar to other *ras* proteins, its biological

functions may differ as R-ras p23 does not share the oncogenic properties of other *ras* proteins^{17,18}.

Clone 1 encodes the 60 C-terminal amino acids of R-ras p23, indicating that the C-terminal hypervariable domain of R-ras p23 is sufficient for interaction with Bcl-2 (Fig. 1b). Clones 2 and 3 encode a larger part of R-ras p23, starting at amino-acid 71 (Fig. 1a). None of the three isolated R-ras clones was able to interact with p53 (Fig. 1b), or with the C-terminal regions of K-ras and H-ras, or with *rac* or *rho* proteins (not shown), indicating that the interaction between Bcl-2 and R-ras p23 is specific and not just the result of a hydrophobic interaction caused by isoprenylation of R-ras p23. Full-length Bcl-2 is evidently necessary for the association with R-ras p23 as the interaction is inefficient with Bcl-2 truncated proteins (Fig. 1c).

To test whether these proteins also interact in human cells, we immunoprecipitated R-ras p23 and Bcl-2 from HeLa and HL60 cell extracts. A GTP-binding assay showed a GTP-binding activity immunoprecipitated with Bcl-2, the size of which could correspond to R-ras p23 (Fig. 2a, lane 2). A GTP-binding protein of the same apparent size immunoprecipitates with R-ras p23 antibodies (Fig. 2b, lane 2). No GTP-binding activity co-immunoprecipitated with p53 (Fig. 2a, lane 1). The association of Bcl-2-R-ras p23 seems to be specific for R-ras p23 and not other members of the *ras* family, because no GTP-binding activity was associated with proteins having a relative molecular mass of less than 23,000 (23K) (Fig. 2a, lane 2). Immunoprecipitates of R-ras p23 and Bcl-2 from the membrane fraction of HL60 cells were probed with anti-Bcl-2 antibody. A band matching the size of Bcl-2 was detected in R-ras p23 immunoprecipitates (Fig. 2d, lane 3) but not in control lanes (Fig. 2d, lanes 4, 5).

An association of a GTP-binding activity with Bcl-2 led previously to the suggestion that Bcl-2 itself was a GTP-binding protein¹⁹, but it was later shown that under more stringent conditions²⁰, Bcl-2 was unable to bind GTP. Our results indicate

FIG. 1 Identification of R-ras as a Bcl-2 binding protein. A full-length Bcl-2 protein is required to associate with R-ras p23. a, Relative position of clones 1, 2 and 3 to the full-length R-ras cDNA. Clone 1 encodes amino acids 158-218; clones 2 and 3 encode amino acids 71-218. b, R-ras p23 specifically interacts with Bcl-2. Left, R-ras clone 1 and GAL4-p53 fusion transformed into YGH1 and plated on medium without histidine. Right, R-ras clone 1 and GAL4-bcl-2 cDNA transformed into YGH1 and plated on medium without histidine. c, Maps of Bcl-2 deletions. Numbers indicate the first and last amino acid of each Bcl-2 deletion. Bcl-2.1 represents full-length Bcl-2. Deletions were fused to the GAL4 DNA-binding domain (DBD GAL4). Relative growth: (++) normal growth; (+) slow growth; (-) no growth. METHODS. a, The cDNA encoding human Bcl-2 (ref. 5) was fused to the GAL4-DNA binding domain in the pGBT8 plasmid, and transformed into the yeast strain YGH1. The YGH1 strain carrying GAL4-Bcl-2 plasmid was transformed with a HeLa cDNA library fused to the GAL4 activation domain in the pGAD plasmid¹⁵. When a cDNA encodes a protein that interacts with the Bcl-2 protein, the YGH1 strain is expected to grow in the absence of histidine and produces β -galactosidase. Twenty of the 2×10^6 transformants screened grew in the absence of histidine and had β -galactosidase activity. The plasmids recovered from these 20 yeast strains were used to retransform the original YGH1 + Bcl-2 strain. All the plasmids conferred the ability to grow in the absence of histidine and to produce β -galactosidase. Three of the plasmids contained cDNAs encoding human R-ras. b, The YGH1 yeast strain transformed with clone 1 (see a), and either *bcl-2* cDNA or p53 cDNA, fused to the GAL4 DNA-binding domain. Transformants were tested for their ability to grow in the absence of histidine. Growth in the absence of histidine is an indication of the interaction between proteins fused to the GAL4 activation domain (R-ras) and the GAL4 DNA-binding domain (Bcl-2 or p53). Plates were incubated at 30 °C for 4 days. For c, deletions were made using standard techniques.

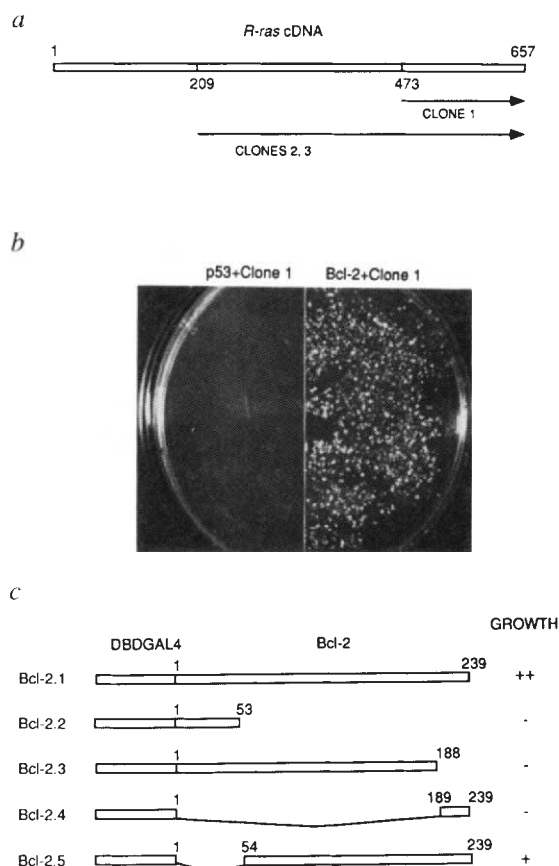
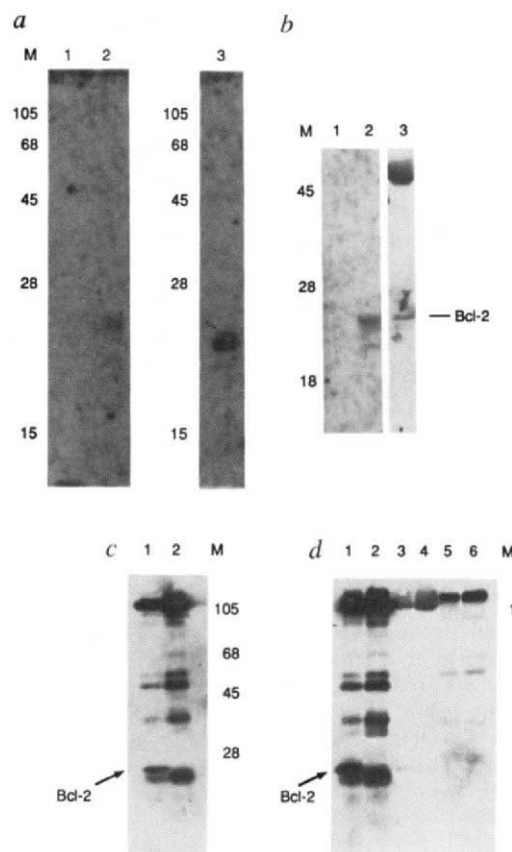


FIG. 2 A 23K GTP-binding protein co-immunoprecipitates with Bcl-2. Bcl-2 co-immunoprecipitates with R-ras p23. **a**, GTP blot of p53 and Bcl-2 immunoprecipitates. Lane 1, HeLa cell lysate immunoprecipitated with a monoclonal antibody against p53; lane 2, HeLa cell lysate immunoprecipitated with a monoclonal antibody against human Bcl-2; lane 3, 100 ng purified bacterial human R-ras protein. **b**, R-ras p23 and Bcl-2 co-migrate during electrophoresis. Lanes 1 and 2, GTP blot of p53 (lane 1) and R-ras p23 (lane 2) immunoprecipitates; lane 3, GTP blot of lane 2 stripped and probed with anti-Bcl-2 antibody. **c** and **d**, Western blots probed with an anti-Bcl-2 monoclonal antibody. Lanes 1 and 2, anti-Bcl-2 immunoprecipitates; lanes 3 and 4, anti-R-ras immunoprecipitates, and lanes 5 and 6, anti-T antigen immunoprecipitates. Lanes 1, 3 and 5 contain HL60 membrane fraction; lanes 2, 4 and 6 contain only lysis buffer. **c**, 5-min exposure of lanes 1 and 2 of the blot; **d**, 30-min exposure of the entire blot. Lanes M, M, markers. **METHODS.** **a**, HeLa cells were washed twice with ice-cold PBS and lysed in 10 mM sodium phosphate (pH 7.4), 150 mM NaCl, 5 mM EDTA, 1% Triton X-100. The lysate was centrifuged at 5,000 r.p.m. in an Eppendorf microfuge at 4 °C and the pellet discarded. Total protein (1 mg) was immunoprecipitated with p53 monoclonal antibody, PAb421 (lane 1) or with a monoclonal antibody against human Bcl-2 (DAKO) (lane 2). As a positive control, 100 ng bacterially produced human R-ras protein was loaded in lane 3. Proteins were resolved on a 11.5% SDS-polyacrylamide gel, transferred to a nitrocellulose filter, and probed with [α -³²P]GTP as described²⁴, except the washing buffer contained 10 mM MgCl₂. **b**, Lysates from HeLa cells were prepared, immunoprecipitated and probed with [α -³²P]GTP. Lane 2 was washed for 1 hour with the GTP-blot wash buffer²⁴ and probed with anti-Bcl-2 antibody as described below. **c**, **d**, HL60 cells were washed twice with ice-cold PBS. Cells were lysed and the P100 membrane fraction was isolated as previously described¹⁷. 1 mg total protein was immunoprecipitated with either a monoclonal antibody against human Bcl-2 (DAKO), rabbit polyclonal antibodies against human R-ras p23 (ref. 16), or monoclonal antibody PAb416 against T antigen. As the Bcl-2 protein co-migrates with the IgG light chain during electrophoresis, we included the above antibodies with no lysate as a control for the western blot (lanes 2, 4 and 6). The immunoprecipitates were washed three times with 50 mM Tris-HCl, pH 7.5, 200 mM NaCl, 0.1% NP-40. Proteins were resolved on a 11.5% SDS-polyacrylamide gel, transferred to a nitrocellulose filter and probed with anti-Bcl-2 antibody. The secondary antibody was peroxidase-conjugated goat anti-mouse IgG, Fc γ fragment-specific (Jackson ImmunoResearch). The ECL system (Amersham) was used to develop the blot.



that the GTP-binding activity¹⁹ was probably due to R-ras p23 associated with Bcl-2. Bcl-2 (M_r 25K) and R-ras (M_r 23K) migrate closely during gel electrophoresis (Fig. 2b, lanes 2 and 3) and this would account for the misinterpretation.

On the basis of our results, R-ras p23 could be a component of a signal transduction pathway that mediates signals (either external or internal) involved in the induction of apoptosis. Perhaps these signals are transmitted to an effector bound to R-ras p23. The association of R-ras p23 with Bcl-2 could prevent this signal from being passed on, resulting in suppression of

apoptosis. Thus, the overexpression of Bcl-2 could alter the ability of R-ras p23 to transmit signals, interrupting the apoptotic signal transduction pathway. Both R-ras p23 and Bcl-2 are localized in the membrane fraction of the cell^{5,8,21,22} and the Bcl-2 membrane localization signal is known to be required for Bcl-2 to inhibit apoptosis²³. This hypothesis could also explain why activated alleles of R-ras are not transforming^{17,18}. A hyperactive R-ras protein would constitutively pass signals for activation of apoptosis, even in the absence of the proper signals. This would result in cell death rather than transformation of the cells. □

Received 17 June; accepted 8 October 1993.

- Nunez, G. et al. *J. Immun.* **144**, 3602–3610 (1990).
- Lee, S., Christakos, S. & Small, M. B. *Curr. Opin. Cell Biol.* **5**, 286–291 (1993).
- Vaux, D. L. *Proc. natn. Acad. Sci. U.S.A.* **90**, 786–789 (1993).
- McDonnell, T. J. et al. *Cell* **57**, 79–88 (1989).
- Hockenbery, D., Nunez, G., Millman, C., Schreiber, R. D. & Korsmeyer, S. J. *Nature* **348**, 334–336 (1990).
- Sentman, C. L., Shutter, J. R., Hockenbery, D., Kanagawa, O. & Korsmeyer, S. J. *Cell* **67**, 879–888 (1991).
- Strasser, A., Harris, A. W. & Cory, S. *Cell* **67**, 889–899 (1991).
- Alnemri, E. S., Robertson, N. M., Fernandes, T. F., Croce, C. M. & Litwack, G. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7295–7299 (1992).
- Bissonnette, R. P., Echeverri, F., Mahboubi, A. & Green, D. R. *Nature* **359**, 552–556 (1992).
- Fanidi, A., Harrington, E. A. & Evan, G. I. *Nature* **359**, 554–556 (1992).
- Garcia, I., Martinou, I., Tsujimoto, Y. & Martinou, J.-C. *Science* **258**, 302–304 (1992).
- Seigel, R. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7003–7007 (1992).
- Vaux, D. L., Weissman, I. L. & Kim, S. K. *Science* **258**, 1955–1957 (1992).
- Wagner, A. J., Small, M. B. & Hay, N. *Molec. cell. Biol.* **13**, 2432–2440 (1993).

- Chien, C.-T., Bartel, P. L., Sternglanz, R. & Fields, S. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9578–9582 (1991).

- Lowe, D. G. et al. *Cell* **48**, 137–146 (1987).
- Lowe, D. G. & Goeddel, D. V. *Molec. cell. Biol.* **7**, 2845–2856 (1987).
- Lowe, D. G., Ricketts, M., Levinson, A. D. & Goeddel, D. V. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1015–1019 (1988).
- Haldar, S., Beatty, C., Tsujimoto, Y. & Croce, C. M. *Nature* **342**, 195–200 (1989).
- Monica, K., Zehava, C.-L. & Cleary, M. L. *Nature* **346**, 189–191 (1990).
- Chen-Levy, Z., Nouse, J. & Cleary, M. L. *Molec. cell. Biol.* **9**, 701–710 (1989).
- Jacobson, M. D. et al. *Nature* **361**, 365–369 (1993).
- Tanaka, S., Saito, K. & Reed, J. C. *J. biol. Chem.* **268**, 10920–10926 (1993).
- Lerosey, I., Chardin, P., de Gunzburg, J. & Tavtitan, A. *J. biol. Chem.* **266**, 4315–4321 (1991).

ACKNOWLEDGEMENTS. We thank D. Lowe of Genentech for anti-R-ras antiserum, S. Korsmeyer for bcl-2 cDNA, G. Hannon for plasmid pGBT8, a HeLa cDNA library and YGH1 yeast strain, G. Torres for help with tissue culture, our co-workers at ONYX for critically reading the manuscript, and especially F.M. for encouragement and for discussion.

Cloning and expression of cytochrome P450 genes controlling flower colour

Timothy A. Holton*, Fillippa Brugliera*,
Diane R. Lester*†, Yoshikazu Tanaka‡,
Craig D. Hyland*, John G.T. Mentling*†,
Chin-Yi Lu*, Ellane Farcy§,
Trevor W. Stevenson*† & Edwina C. Cornish*

* Calgene Pacific Pty Ltd, 16 Gipps Street, Collingwood, Victoria 3066, Australia

‡ Suntory Ltd, 1-1-1 Wadayamadai, Shimamoto-Cho, Mishima-Gun, Osaka 618, Japan

§ INRA, Station de génétique et d'amélioration des plantes, BV 1540, 21034 Dijon cédex, France

† Present addresses: CSIRO Postharvest Laboratory, North Ryde, New South Wales 2113, Australia (D.R.L.); La Trobe University, Bundoora, Victoria 3083, Australia (J.G.T.M.); Royal Melbourne Institute of Technology, Melbourne 3001, Australia (T.W.S.).

BLUE and violet flowers generally contain derivatives of delphinidin; red and pink flowers generally contain derivatives of cyanidin or pelargonidin¹. Differences in hydroxylation patterns of these three major classes of anthocyanidins are controlled by the cytochrome P450 enzymes flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase. Here we report on the isolation of complementary DNA clones of two different flavonoid 3',5'-hydroxylase genes that are expressed in petunia flowers. Restriction-fragment length polymorphism mapping and complementation of mutant petunia lines showed that the flavonoid 3',5'-hydroxylase genes correspond to the genetic loci *Hf1* and *Hf2*.

Anthocyanins are glycosylated derivatives of anthocyanidins, which differ in the degree of hydroxylation (Fig. 1) and subsequent methylation of the B-ring. The addition of extra hydroxyl groups to the anthocyanin B-ring generally leads to a bluing of flower colour¹. Anthocyanin biosynthesis is controlled by several

soluble and microsomal enzymes. Genes encoding many of the soluble enzymes have been cloned by transposon tagging, differential screening and screening with antibodies raised against purified proteins². But genes encoding the microsomal enzymes flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase (F3',5'H) have eluded these cloning strategies.

F3',5'H catalyses the 3',5'-hydroxylation of naringenin and eriodictyol to form 5,7,3',4',5'-pentahydroxyflavanone and 3',5'-hydroxylation of dihydrokaempferol and dihydroquercetin to form dihydromyricetin. F3',5'H activity was first detected in microsomal preparations from flowers of *Verbena*, which contain anthocyanins based on delphinidin³. F3',5'H activity has also been demonstrated on microsomal preparations from flowers of petunia¹. Two genetic loci, *Hf1* and *Hf2*, control F3',5'H activity in flowers of *P. hybrida*⁴. *Hf1* controls activity in both the limb and the tube of flowers and gives rise to much higher F3',5'H activity than *Hf2*, which controls F3',5'H activity only in the limb. F3',5'H activity is also developmentally regulated in petunia flowers (Fig. 2). Subcellular localization, co-factor requirements and inhibitor studies⁵ have suggested that the petunia F3',5'H is a microsomal P450 enzyme.

Isolation strategies for P450 genes have generally relied on the ability to purify the protein and assay for enzyme activity. Purification of P450 enzymes from plants is difficult because of their low abundance⁶ and instability^{7,8}. Although many reactions are catalysed by P450 enzymes in plants, only one plant P450 gene of known function has been cloned⁹. To isolate F3',5'H clones from petunia we used a polymerase chain reaction (PCR)-based strategy. Degenerate oligonucleotides designed to the conserved P450 haem-binding domain were used in PCR-reactions using DNA from a petunia petal complementary DNA library as template. PCR products were cloned¹⁰ and those encoding putative P450 sequences were used to screen a petunia cDNA library for related sequences. The combined PCR amplification and screening identified cDNA clones of 18 different genes expressed in petunia petals (data not shown). The molecular analysis of two petunia P450 cDNA clones, pCGP175 and pCGP176, encoding F3',5'H, are described.

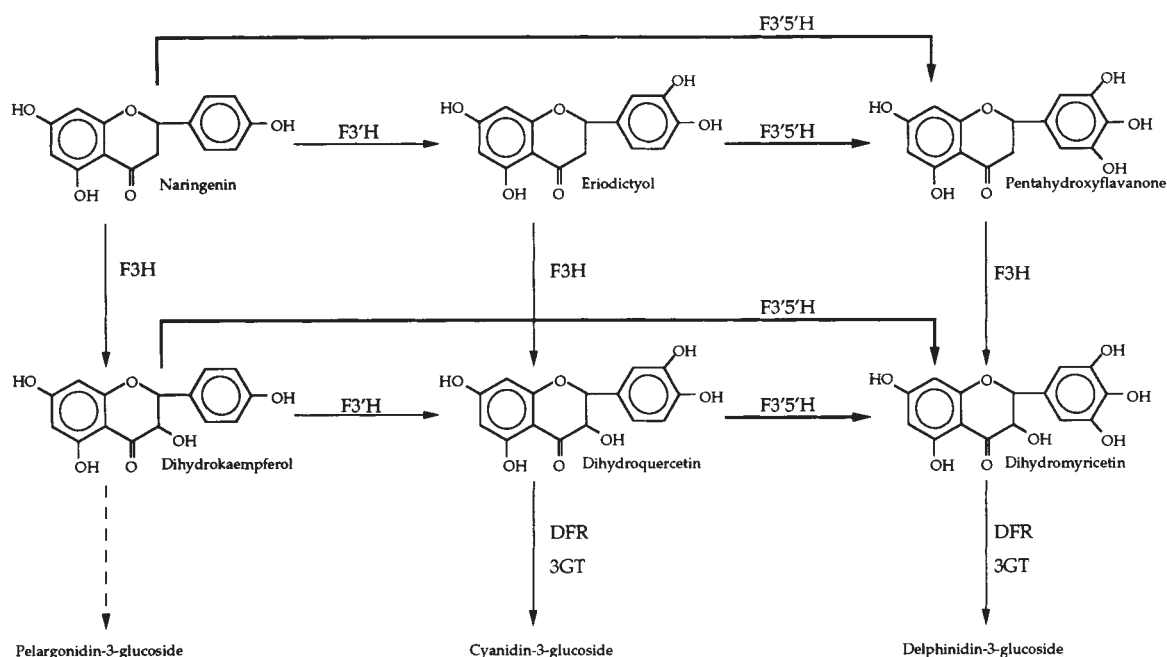


FIG. 1 Anthocyanin biosynthesis pathway. Enzymes involved in the synthesis of anthocyanidin 3-glucosides are indicated as follows: F3H, flavanone 3-hydroxylase; F3'H, flavonoid 3'-hydroxylase; F3',5'H, flavonoid 3',5'-hydroxylase; DFR, dihydroflavonol-4-reductase; 3GT,

UDP-glucose:flavonoid 3-O-glucosyltransferase. In petunia, pelargonidin derivatives are not usually produced owing to the substrate specificity of DFR. Further modifications of cyanidin and delphinidin 3-glucosides are possible depending on the genotype.

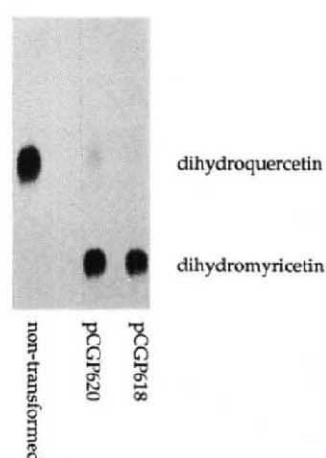
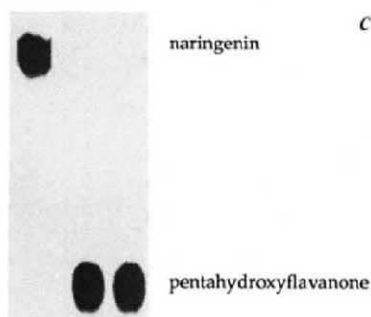
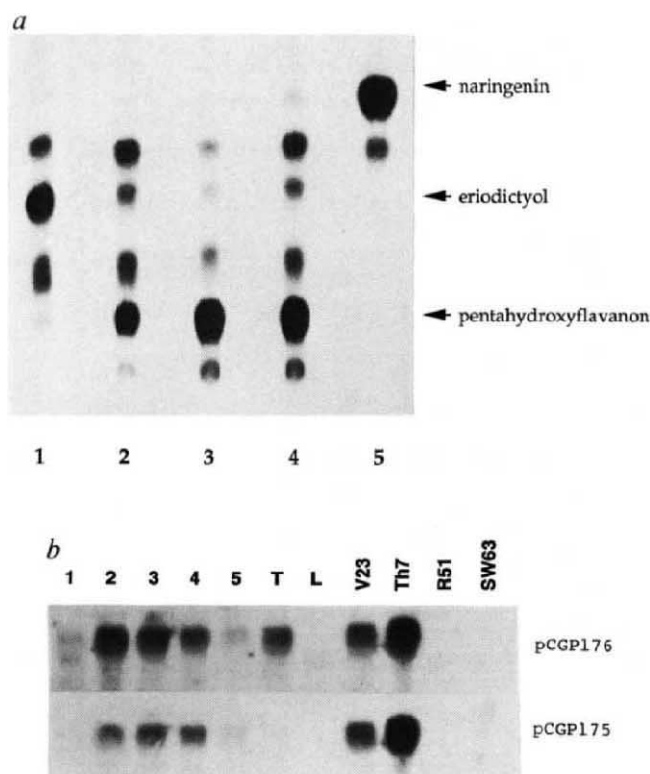


FIG. 2 Identification of F3'5'H cDNA clones. **a**, Five developmental stages of *Petunia hybrida* cv. Old Glory Blue (OGB) flowers (1–5)¹⁷ were assayed for F3'5'H activity by conversion of naringenin to 5,7,3',4',5'-pentahydroxyflavanone using ³H-labelled naringenin. **b**, Northern blot analysis of RNA from: 1–5, petal limbs of OGB flowers from the five different stages of flower development; T, petal tube tissue from stage 3–4 OGB flowers; L, leaf tissue from OGB seedlings; V23, Th7, R51 and SW63, petal limbs of *P. hybrida* cultivars. The cDNA clones pCGP175 and pCGP176 were used as probes. The cDNA fragments from pCGP175 and pCGP176 were cloned into a yeast expression vector to produce pCGP618 and pCGP620, respectively. The yeast strain G-1315 (ref. 15) was transformed¹⁶ with pCGP618 and pCGP620. Microsomal fractions from yeast transformants and from control (untransformed) yeast were assayed for F3'5'H activity using *c*, [³H]-naringenin, and *d*, [³H]-dihydroquercetin as substrates. METHODS. Plant tissue from five defined stages of flower development¹⁷ was assayed for F3'5'H activity using [³H]-naringenin as substrate⁵. RNA was isolated¹⁸ from petunia tissues. RNA samples (20 µg) were electrophoresed through a 2.2 M formaldehyde/1.2% (w/v) agarose gel, transferred to Hybond-N membrane (Amersham) and probed with ³²P-labelled cDNA fragments of pCGP175 or pCGP176. A 1.8-kb *EcoRI*/*KpnI* fragment containing the entire cDNA insert from pCGP175 was ligated with a 9-kb *EcoRI*/*KpnI* fragment from the yeast expression vector pYGA2269 (ref. 15). The resulting plasmid, pCGP618, contained the pCGP175 cDNA fragment in a sense orientation behind the glyceraldehyde-3-phosphate promoter. The pCGP176 cDNA was similarly cloned into pYGA2269 to produce pCGP620. Single isolates of G-1315 yeast transformed with pCGP618 or pCGP620 were grown in liquid culture for 2 d at 30 °C¹⁹. Microsomal pellets were isolated, suspended in 400 µl of buffer A¹⁹ and a 100 µl aliquot was assayed for F3'5'H activity.

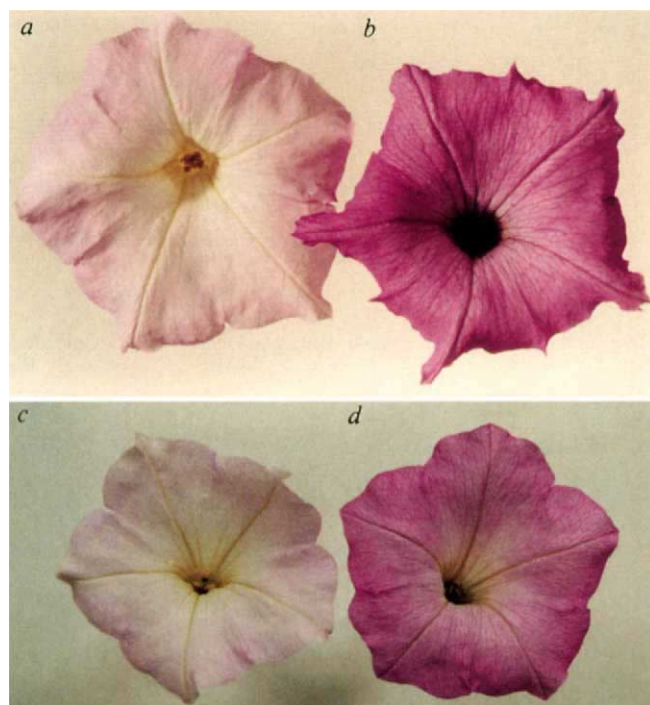


FIG. 3 Complementation of F3'5'H mutations in petunia. The F3'5'H cDNA inserts from pCGP602 and pCGP175 were cloned into a binary expression vector to create pCGP90 and pCGP802, respectively. *P. hybrida* cv. Skr4 x Sw63 (*hf1hf1hf2hf2*) was transformed with pCGP90 and pCGP802. **a** and **c**, Flowers from non-transformed Skr4 x Sw63. **b**, Flower from Skr4 x Sw63 transformed with pCGP90. **d**, Flower from Skr4 x Sw63 transformed with pCGP802.

METHODS. A 1.8-kb *Bam*HI/*Kpn*I fragment containing the cDNA from pCGP602 (includes 125 bp upstream of the initiation codon) was cloned in a sense orientation between the *Mac* promoter and *mas* terminator of the binary vector pCGP293 (ref. 17), to create pCGP90. A 1.8-kb *Bam*HI/*Kpn*I fragment containing the cDNA from pCGP175 (*Hf2*) was cloned in a sense orientation between the *Mac* promoter and *mas* terminator of pCGP293 to create pCGP802. Gene constructs were transferred to petunia via co-cultivation of leaf disks²⁰ with *Agrobacterium tumefaciens* AGLO²¹ containing the binary plasmids pCGP90 or pCGP802.

a

1 CTTTCTACTAGCTACTTCGTTATATATATGTAATAATGTGACTTTGAAAAATCATTTAAATATCATAGGTTCAATTTATCTTGATCAAAATATTACTTCGGCCATATACGTTTTCCTT
 121 TAGTCATGATGCTACTTACTGAGCTTGGTGCAGCAACTTCAATCTTTCTAATAGCACACATATCAATTTCAACTCTTATTTTCAAAATACCAGCCGGCATCTACCGCCGGGGCAAGAG
 M M L L T E L G A A T S I F L I A H I I I S T L I S K T T G R H L P P G P R
 241 GGTGGCCGGTGATCGGAGCACTTCCACTTTTAGGAGCCATGCCACATGTTTCCTTAGCTAAATGCAAAAAATATGGAGCAATCATGTATCTCAAAGTTGGAACATGTGGCATGGCAG
 G W P V I G A L P L L G A M P H V S L A K M A K K Y G A I M Y L K V G T C G M A
 361 TTGCTTCTACCCCTGATGCTGCTAAAGCATTTCTGAAAACACTTGATATCAACTTCTCAATCGTCCACCTAATGCAGGTGCCACTCACTAGCTTATAATGCTCAAGACATGGTTTTG
 V A S T P D A A K A F L K T L D I N F S N R P P N A G A T H L A Y N A Q D M V F
 481 CACATTATGGACCAGTGAAGTTGCTAAGGAAATTAAGCAACTTGCATATGCTAGGGGGAAGCCCTTAGAGAATTGGGCAATGTTGCTGCCAATGAGCTAGGGCAGATGCTAAAAAT
 A H Y G P R W K L L R K L S N L H M L G G K A L E N W A N V R A N E L G H M L K
 601 CAATGTCGATATGAGTCGAGAGGGCCAGAGGTTGTGGTGGCGAGATGTTGACATTTGCCATGGCCAAATATGATCGGACAAGTATGCTAAGCAAAAGAGTATTTGTAGATAAAGGTG
 S M S D M S R E G Q R V V V A E M L T F A M A N M I G Q V M L S K R V F V D K G
 721 TTGAGGTAAATGAATTTAAGACATGGTTGATGAGTTAATGACAATAGCAGGATTTTCAACATTTGGTGATTTTATCTTGTGTTAGCTTGGATGGATTACAAGGGATAGAAAAACGAA
 V E V N E F K D M V V E L M T I A G Y F N I G D F I P C L A W M D L Q G I E K R
 841 TGAACGTTTACATAAGAAGTTTGATGCTTTTATGACAAAGATGTTTGATGAACACAAAGCAACTACCTATGAACGTAAGGGGAACAGATTTTCTTGATGTTGTTATGGAATAAGGGG
 M K R L H K K F D A L L T K M F D E H K A T T Y E R K G K P D F L D V V M E N G
 961 ACAATCTGAAGGAGAAAGACTCAGTACAACCAACATCAAGCACTTTTGCTGAATTTGTTTCAAGCTGGTACGAGCACTTCTTCTAGTGAATAGAATGGGCACTTGCAAGAAATGATGA
 D N S E G E R L S T T N I K A L L L N L F T A G T D T S S S A I E W A L A E M M
 1081 AGAACCTGCCATTTTGAAGAAAGCACAAGCAAGATGAAGTCAATGGAAGAAATAGGCGTTTACTCGAATCGGATATCCCAATCTCCCTTACCTCCGAGCAATTTGCAAGAAA
 K N P A I L K K A Q A E M D Q V I G R N R L L E S D I P N L P Y L R A I C K E
 1201 CATTTGAAACACCCCTTACACCATTAATCTTCTAGGATCTCGAACGAACCATGATAGTGCATGTTTATACATACCAAAAAACACTAGGCTTGTGTTAATATGGCAATG
 T F R K H P S T P L N L P R I S N E P C I V D G Y Y I P K N T R L S V N I W A I
 1321 GAAGAGATCCCCAAGTTTGGGAAAATCCACTAGAGTTTAATCCGAAAGATCTTGAGTGAAGAAATCCAAAGATTGATCCTCGAGGGAACGATTTTGAATGATACCATTTGGTGCTG
 G R D P Q V W E N P L E F N P E R F L S G R N S K I D P R G N D F E L I P F G A
 1441 GACGAAGAATTTGTCAGGAACAAAGATGGAATTTGAATGGTGAATATATATAGGAACCTTTGGTTCATTGTTGATTGGAATTTACCAAGTGAAGTTATGAGTTGAATATGGAAG
 G R R I C A G T R M G I V M V E Y I L G T L V H S F D W K L P L S E V I E L N M E
 1561 AAGCTTTTGGCTTAGCTTTGAGAAAGCTTCCCTCTGAAGCTATGGTTACTCCAAGTTTACAATTTGATGTTTATGTACCATAGCTATAGATGTGATTGTGCTATAATTGCGCATGT
 E A F G L A L Q K A V P L E A M V T P R L Q L D V Y V P *
 1681 TGTGTGTTGATGATGATATTAAGAGGATACATGAAGCGCATTCGATGAGTTTAACTTGTAGCTCCTTAATATTTTAGGTATTTTCAATTAATAAGTCTTGTGTGGTGGGTATT
 1801 TTACAGAATTTAGTACTATTATTTGTCAATTTAGAATTTGTACGCTGAATTTGTGTTCGCCCTTTTATTTTCACTTTGATTGTATCCAAAGGATGCGAATGAATCATACTCTTTA
 1921 CCT

b

1 TTGAATCCAGCTCTATCTGGCTTTAGACAATGGTCTACTTAGTGAGCTTGTGTCAGCAACCTTAATCTTTCTAACAACATATCTTCATTTCAACTCTTCTTTCTATAACTAACGGCC
 M V L L S E L A A A T L I F L T T H I F I S T L S I T N G
 121 GCGCTCTCCGCCAGGGCCAGAGGGTGGCCGGTGATCGGAGCACTTCCACTTTTAGGAGCCATGCCACATGTTTCCTTAGCTAAATGCAAAAAATATGGAGCAATCATGTATCTCA
 R R L P P G P R G W P V I G A L P L L G A M P H V S L A K M A K K Y G A I M Y L
 241 AAGTTGGAACGTTGGCATGGTAGTTGCTTCTACCCCTGATGCTGCTAAAGCTTCTGAAAACACTTGATCTCAACTTCTCAATCGTCCACCTAATGCAGGTGCCACCCACTTAGCCT
 K V G T C G M V V A S T P D A A K A F L K T L D L N F S N R P P N A G A T H L A
 361 ATGGTGCTCAAGACATGTTTGTGCACATATGGACCAAGATGGAAGTTGCTAAGGAAATTAAGCAACTTACATATGCTAGGGGGAAGCCCTTAGAAAATTTGGGCAATGTTGCTGCCA
 Y G A Q D M V F A H Y G P R W K L L R K L S N L H M L G G K A L E N W A N V R A
 481 ATGAGCTAGGACACATGCTAAATCGATGTTTGATATGAGCAGAGAAGGGAGAGAGTTGTGGTGGCGAGATGTTGACATTTGCCATGGCAATATGATCGGACAGGTGATAGTCTAGCA
 N E L G H M L K S M F D M S R E G E R V V V A E M L T F A M A N M I G Q V I L S
 601 AAAGAGTATTTGTAATAAGGTTGTTGAGTAAATGAATTTAAGGACATGGTGGTAGAGTTAATGACAACAGCAGGATTTTAACTTTGGTGATTTTATCTTGTGTTAGCTTGGATGG
 K R V F V N K G V E V N E F K D M V V E L M T T A G Y F N I G D F I P C L A W M
 721 ATTTCAAGGGATAGAAAAAGGAATGAACGTTTACATAAGAAGTTTGATGCTTTTATGACAAAGATGTTTGATGAACACAAAGCAACTAGCTATGAACGTAAAGGGAAACAGATTTTC
 D L Q G I E K M K R L H K K F D A L L T L T K M F D E H K A T T Y E R K G K P D F
 841 TTGATTGTGTTATGAAAATAGGGCAATTTGAGGAGAAAGGCTCAGTACAACCAACATCAAGCACTTCTGCTGAATTTGTTTCAAGCTGGTACAGACACTTCTTCTAGTGAATAG
 L D C V M E N R D N S E G E R L S T T N I K A L L L N L F T A G T D T S S S A I
 961 AATGGGCACTTGCAGAGATGATGAAGAACCTGCCATTTTAAAGAAAGCACAAGGAGAAATGAAGTCAATGGAAGAAATAGGCGTCTGCTGAATCGGATATCCCAATCTCCCTT
 E W C A L A E M M K N P A I L K K A Q G E M D Q V I G N N R R L L E S V I E L N M E
 1081 ACCTCCGAGCAATTTGCAAGAAACATTTGCAAGCACCTTACACCATTAATCTCCCTAGGATCTCGAACGAACCATGCAATGTCGATGTTTATACATACCAAAAAACACTAGGC
 Y L R A I C K E T F R K H P S T P L N L P R I S N E P C I V D G Y Y I P K N T R
 1201 TTAGTGTAAACATATGGCAATTTGAAGAGATCCCGAAGTTTGGGAGAACCCACTAGAGTTTATCTGAAAGGTTCTTGAGTGAAGAAATCGAAGATTGATCCTCGAGGGAACGACT
 L S V N I W A I G R D P E V W E N P L E F Y P E R F L S G R N S K I D P R G N D
 1321 TTGAATTTGATACCATTTGGTGTGGACGAAGAATTTGTGAGGACAAAGATGGAATCGTAATGGTGAATATATATAGGAACCTTTGGTCCATTTGATTGGAATTTACCAAGTG
 F E A L I P F G A R R I C A G T R M G I V M V E Y I L G T L V H S F D W K L P L S
 1441 AAGTTATGAGCTAAATATGAAGAAGCTTTGGATAGCTTTGAGAAAGCTTCCCTCTGAAGCTATGGTTACTCCAAGGCTGCCTATTGATGTTTATGACCTTTAGCTTGAAACA
 E V I E L N M E E A F G L A L Q K A V P L E A M V T P R L P I D V Y A P L A *
 1561 TGCCTTTACGTTGGTTTCACTTTGGGTAGTATAATGTTGGTGTGCTATAGAAATATTAATAAATGCTAGTATCTTGAAGCGCGTGCAGGGGAGGGGTTGCTCTAGATAGTA
 1681 GTAATATGTTAGCTTCTCTTTATTTCTTGTGATGTTGAGAATCTGATATGTTTCTG

FIG. 4 Nucleotide and predicted amino-acid sequence of *a*, the pCGP602/pCGP176 cDNAs and *b*, the pCGP175 cDNA. Both strands of the cDNA inserts were completely sequenced using Sequenase (US Biochemical). The sequence was compiled from subclones and from sequencing reactions using gene specific oligonucleotide primers. The amino-acid sequence derived by conceptual translation is shown; stop

codons are indicated by asterisks. The sequences of the cDNA inserts from pCGP176 and pCGP602 were identical except that the pCGP602 cDNA started at position 1 and ended at position 1796 and the pCGP176 cDNA started at position 98 and ended at position 1923. The sequences of the petunia F3'5'H cDNA clones have been submitted to the EMBL database with the accession numbers Z22544 and Z22545.

Northern analysis of petunia petal RNA using a pCGP176 cDNA probe revealed a 1.8-kilobase (kb) transcript present in the flowers of petunia cultivar Old Glory Blue (OGB) and in genetically defined *Hf1Hf2* petunia lines but not in *hf1hf1hf2hf2* lines (Fig. 2). The gene corresponding to pCGP176 was expressed in both the tube and the limb of OGB flowers but was not expressed in leaves. A pCGP175 cDNA probe also detected a 1.8-kb transcript which was expressed in the limb and at much lower levels in the tube of OGB flowers but was not expressed in *hf1hf1hf2hf2* flowers (Fig. 2). Maximum levels of pCGP175 and pCGP176 transcripts correlated with the flower stages producing maximum F3'5'H activity. Therefore the genes corresponding to the pCGP176 and pCGP175 cDNA clones had expression patterns that might be expected of *Hf1* and *Hf2*, respectively.

Hf1 is located on chromosome I whereas *Hf2* is closely linked to *Po* on chromosome V (ref. 11). Restriction-fragment-length polymorphism (RFLP) analysis of DNA isolated from an F₂ population of plants from a cross between two inbred petunia lines, V23 (*Hf1Hf1Hf2Hf2*) and R51 (*hf1hf1hf2hf2*), was used to obtain linkage data¹² for the P450 genes. An *Hf1*− genotype was assigned to F₂ plants that expressed F3'5'H activity in the flower tube. All of the 49 *Hf1*− plants analysed produced a V23-like or heterozygous pattern when probed with the pCGP176 cDNA, revealing a close linkage to *Hf1*. A pCGP175 cDNA probe detected an RFLP that showed complete segregation with *Po* in 32 plants, indicating that the gene corresponding to pCGP175 was closely linked to *Po* and therefore to *Hf2*.

To determine whether pCGP175 or pCGP176 encoded F3'5'H, the cDNA inserts were cloned into a yeast expression vector to produce pCGP618 and pCGP620, respectively. F3'5'H activity was detected in yeast transformed with either pCGP618 or pCGP620 but not in non-transformed yeast (Fig. 2). These results demonstrated that the cDNA inserts from pCGP175 and pCGP176 encoded F3'5'H enzymes able to 3',5'-hydroxylate naringenin and dihydroquercetin.

The F3'5'H cDNA inserts from pCGP602 (a longer version of pCGP176 containing 125 base pairs upstream of the initiation codon) and pCGP175 were cloned into a binary expression vector to produce pCGP90 and pCGP802, respectively. The petunia hybrid *Skr4* × *Sw63* (*hf1hf1hf2hf2*) was transformed with pCGP90 or pCGP802 and transgenic plants were grown to flowering (Fig. 3). Flowers of the transgenic plants produced higher levels of anthocyanins and had significantly higher levels of both F3'5'H activity and 3',4',5'-hydroxylated anthocyanins than the non-transgenic plants (data not shown). The combination of RFLP mapping and complementation results suggested that both *Hf1* and *Hf2* encode structural genes for F3'5'H enzymes.

Nucleotide sequences of the cDNA inserts of pCGP175, pCGP176 and pCGP602 were determined (Fig. 4). The pCGP602 (*Hf1*) and pCGP175 (*Hf2*) cDNA sequences are 93% identical in nucleotides and 94% identical in amino acids. Comparison of the F3'5'H sequences with the EMBL and Genbank databases revealed that the most similar sequence was an avocado P450 (ref. 13) which shared ~38% amino-acid identity with both petunia F3'5'H genes. On the basis of the P450 gene classification system¹⁴, both F3'5'H genes belong to a new P450 family. □

Received 14 May; accepted 9 September 1993.

1. Forkmann, G. *Pl. Breeding* **106**, 1–26 (1991).
2. Dooner, H. K. & Robbins, T. P. *A. Rev. Genet.* **25**, 173–199 (1991).
3. Stotz, G. & Forkmann, G. *Z. Naturforsch.* **37c**, 19–23 (1982).
4. Wiering, H. *Genen. Phaenen* **17**, 117–134 (1974).
5. Menting, J. G. T. thesis, La Trobe Univ. (1992).
6. Riviere, J.-L. & Cabanne, F. *Biochimie* **69**, 743–752 (1987).
7. Stewart, C. B. & Schuler, M. A. *Pl. Physiol.* **90**, 534–541 (1989).
8. Jarvis, L. & Pierpoint, W. S. *J. Biotechnol.* **11**, 161–198 (1989).
9. Mitzutani, M. et al. *Biochem. biophys. Res. Commun.* **190**, 875–880 (1993).
10. Holton, T. A. & Graham, M. W. *Nucleic Acids Res.* **19**, 1156 (1991).
11. Cornu, A. et al. in *Genetic Maps—Locus Maps of Complex Genomes* 5th edn (ed. O'Brien, S. J.), (Cold Spring Harbor Laboratory Press, New York, 1990).

12. Wallroth, M., Gerats, A. G. M., Rogers, S. G., Fraley, R. T. & Horsch, R. B. *Molec. gen. Genet.* **202**, 6–15 (1986).
13. Bozak, K. R., Yu, H., Sirevag, R. & Christoffersen, R. E. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3904–3908 (1990).
14. Nebert, D. W. et al. *DNA Cell Biol.* **10**, 1–14 (1991).
15. Ashikari, T. et al. *Appl. Microbiol. Biotechnol.* **30**, 515–520 (1989).
16. Ito, H., Fukuda, Y., Murata, K. & Kimura, A. *J. Bact.* **153**, 163–168 (1983).
17. Brugiera, F. et al. *Pl. J.* (in the press).
18. Turpen, T. H. & Griffith, O. M. *BioTechniques* **4**, 11–15 (1986).
19. Oeda, K. & Kakaki, T. *DNA* **4**, 203–210 (1985).
20. Horsch, R. B. et al. *Science* **227**, 1229–1231 (1985).
21. Lazo, G. R., Pascal, A. S. & Ludwig, R. A. *Bio/Technology* **9**, 963–967 (1991).

ACKNOWLEDGEMENTS. We thank C. Caesar, S. Tsuda and I. Gautrais for petunia transformations, A. Simmons for maintenance of transgenic plants, T. Perilli for technical assistance and M. Michael for help with sequencing P450 cDNA clones.

Identification of nucleotide-binding regions in the chaperonin proteins GroEL and GroES

Jörg Martin, Scott Geromanos*, Paul Tempst* & F. Ulrich Hartl†

Cellular Biochemistry and Biophysics Program, and * Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York 10021, USA

THE chaperonin GroEL, a tetradecameric cylinder consisting of subunits of *M_r* ~60,000 (60K), and its cofactor GroES, a heptameric ring of 10K subunits, mediate protein folding in the cytosol of *Escherichia coli*^{1–3}. In the presence of nucleotide, GroES forms a 1:1 complex with GroEL which binds unfolded protein in its central cavity and releases it to allow folding upon ATP hydrolysis^{4–7}. Using labelling with azido-ATP, we have identified a protease-stable nucleotide-binding domain of *M_r* 40K in the GroEL subunits (residues 153–531). Azido-ATP is crosslinked to the highly conserved Tyr 477, indicating that this residue is close to the purine ring of the bound nucleotide. Surprisingly, GroES also binds ATP cooperatively and with an affinity comparable to that of GroEL. Azido-nucleotide labelling of GroES subunits occurs at the conserved Tyr 71 in a protease-stable 6.5K domain (starting at residue 33). Proteinase K cleavage at residue 32 is prevented when GroES is bound to GroEL. ATP binding to GroES may be important in charging the seven subunits of the interacting GroEL ring with ATP to facilitate cooperative ATP binding and hydrolysis for substrate protein release.

The double-toroid GroEL protein has 14 ATP binding sites⁸. Binding of GroES stabilizes the interacting seven-subunit ring of GroEL in a high-affinity state for ADP and also favours ADP binding to the opposite toroid⁹. This results in an inhibition of the GroEL ATPase^{10–12}. Binding of unfolded protein leads to the dissociation of ADP and, as a consequence, to GroES release⁹. GroES rebinds upon exchange of ADP with ATP⁸, and subsequent cooperative ATP hydrolysis in GroEL^{13,14} discharges the substrate protein for folding. For the identification of the nucleotide-binding region in the GroEL subunits, purified GroEL was incubated with [α -³²P]8N₃-ATP or [α -³²P]8N₃-ADP and bound nucleotide was covalently attached by ultraviolet irradiation (Fig. 1A). GroEL hydrolyses 8N₃-ATP at ~70% of the ATP rate (not shown). Covalent labelling was dependent on Mg²⁺ and was significantly enhanced by complex formation with GroES, reflecting the GroES-induced increase in the nucleotide binding affinity of GroEL^{8,13}. Unexpectedly, GroES itself was found to be crosslinked to 8N₃-ATP and 8N₃-ADP in these experiments (Fig. 1A), although labelling with 8N₃-ADP was less efficient.

† To whom correspondence should be addressed.

The binding of 8N₃-ATP to GroES and covalent labelling was competed by ATP (Fig. 1*B,a*). Interestingly, the efficiency of 8N₃-ATP crosslinking initially increased at low concentrations of ATP, an effect that was also observed with nucleotide binding to GroEL. To analyse whether ATP binding is indeed cooperative, the chaperone proteins were incubated at 4 °C or 23 °C with increasing concentrations of ATP containing 0.6% [α -³²P]8N₃-ATP followed by ultraviolet crosslinking (Fig. 1*B,b* and *c*). Labelling of GroES was saturable at both temperatures. At 23 °C, a cooperative binding curve with a Hill coefficient of 2.0 was obtained. In a separate experiment, at least 3–4 molecules of 8N₃-ATP could be attached per GroES oligomer by subsequent rounds of crosslinking, as visualized by a shift of the labelled subunits on SDS-PAGE (not shown). It is therefore reasonable to assume that the heptameric GroES has seven ATP binding sites. Half-maximal labelling with 8N₃-ATP was observed at ~300 μ M ATP (Fig. 1*B,b*). The affinity of GroES for 8N₃-ATP may be lower than that for ATP, however. The same consideration applies to GroEL which bound 8N₃-ATP with an apparent affinity comparable to GroES (Fig. 1*B,c*). Previous attempts to isolate GroEL with bound ATP or ADP have failed owing to the low nucleotide affinity of GroEL in the absence of GroES^{8,9}. Using crosslinking of 8N₃-nucleotide as the measure, we have detected a cooperativity in ATP binding to GroEL alone. Hill coefficients of 2.2 at 23 °C and 1.6 at 4 °C were found.

Nucleotide-binding sites in GroEL and GroES have not yet been identified. By covalent labelling with [α -³²P]8N₃-ATP, we were able to identify specific amino-acid residues of GroEL and

GroES that are in close proximity to the bound nucleotide. The analysis of GroEL and GroES as a complex resulted in more efficient labelling of GroEL in these experiments. Labelled protein was digested with trypsin and then applied onto an aluminium chelate column which binds phosphopeptides¹⁵. Bound peptides were eluted and separated by reversed-phase HPLC (Fig. 2*a*). The covalent coupling of 8N₃-nucleotide is known to have a strong impact on the fractionation pattern of peptides under these conditions. Moreover, the 8N₃-nucleotide group is unstable upon prolonged exposure at low pH during HPLC, giving rise to breakdown products¹⁵. Consistent with this, several incompletely separated peptide peaks with an absorbance at both 214 nm and 297 nm were detected containing radioactivity. As GroEL and GroES lack tryptophan residues, the absorbance at 297 nm at pH 2.0 is indicative of 8N₃-nucleotide labelling. The two principal sections of the HPLC profile were pooled separately as indicated in Fig. 2, and were analysed by amino-acid sequencing. Two peptide sequences, one of GroEL (residues 470–497) and one of GroES (residues 61–74), were identified (Fig. 3*a*). The same two sequences were detected in both pooled fractions. As expected, mass spectrometry demonstrated heterogeneity among the peptides of identical sequence with relative molecular mass differences of 160–500 (not shown). Significantly, Tyr 477 in GroEL and Tyr 71 in GroES were the only residues not detectable in these peptides upon sequencing, indicating that they contained the attached 8N₃-nucleotide. In certain ATP-binding proteins, tyrosine is found close to the purine ring^{16,17}. Comparison of known GroEL and GroES sequences

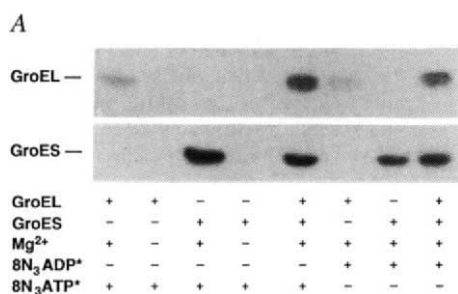


FIG. 1 A, Covalent crosslinking of 8N₃-nucleotides to GroEL and GroES. B, a, Competition of 8N₃-ATP crosslinking by unlabelled ATP, and saturation of 8N₃-ATP binding and crosslinking to GroES (b) and GroEL (c) at 4 °C (●) and 23 °C (○).

METHODS. GroEL and GroES proteins were purified from an overproducing strain of *E. coli* bearing the plasmid pOF39 (refs 4, 21). [α -³²P]8N₃-ADP was prepared by incubation of the ATP form (9.47 mCi μ mol⁻¹; ICN) with 5 μ g ml⁻¹ hexokinase in the presence of 10 mM glucose. Nucleotide was separated from hexokinase by centrifugation through Microcon-10 (Amicon). Conversion was complete, as tested by thin-layer chromatography²². A, GroEL (120 nM 14-mer) or GroES (100 nM 7-mer) were incubated in separate reactions or together with a mixture of 5 μ M [α -³²P]8N₃-ATP/80 μ M ATP or 5 μ M [α -³²P]8N₃-ADP/80 μ M ADP in buffer A (20 mM Tris, pH 7.4, 100 mM KCl, 5 mM magnesium acetate); where indicated, magnesium acetate was omitted. Reactions were incubated at 23 °C for 15 min and then irradiated for 30 s with a 312-nm UV lamp (FB-UVM-80; Fisher) ($I_{15\text{ cm}} = 860 \mu\text{W cm}^{-2}$) at a distance of 1.5 cm. Proteins were analysed by SDS-PAGE and fluorography. The lower panel was exposed to X-ray film four times longer than the upper one. B, a, GroEL (30 nM) or GroES (60 nM) were incubated for 1 min at 23 °C in buffer A containing 1 μ M [α -³²P]8N₃-ATP. Increasing concentrations of ATP were added for 1 min before UV crosslinking. After analysis by SDS-PAGE, fluorographs were quantified by laser densitometry. b, GroES (30 nM), or c, GroEL (60 nM) were incubated for 1 min at 4 °C or 23 °C in buffer A with the indicated total concentrations of a mixture of 50 μ M [α -³²P]8N₃-ATP/8 mM ATP. After UV irradiation reactions were analysed by SDS-PAGE and fluorography. Hill coefficients were obtained after transformation of the quantified data into a linearized Hill plot (not shown).

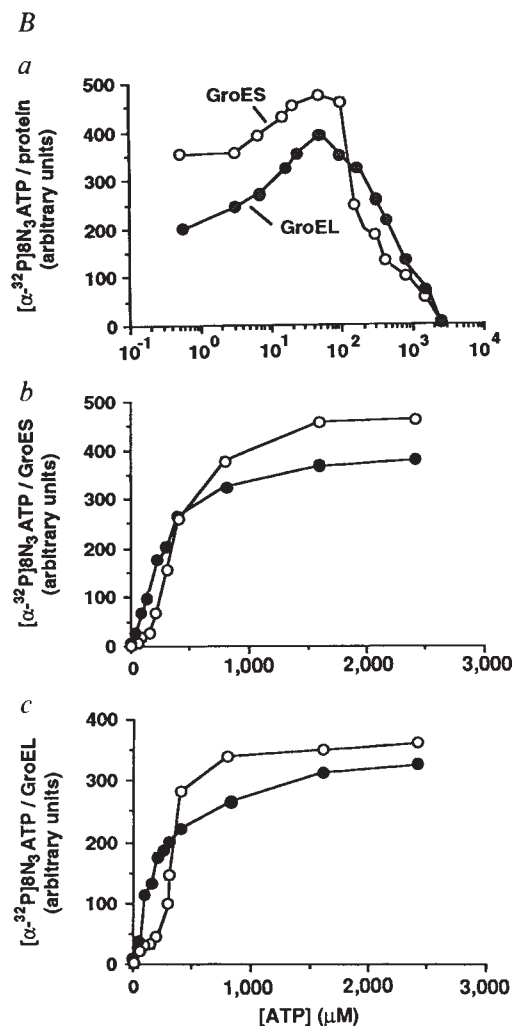


FIG. 2 Identification of $8N_3$ -ATP-modified amino acid residues in GroEL and GroES. Fractionation by reversed-phase HPLC of [α - ^{32}P] $8N_3$ -ATP-labelled tryptic peptides of GroEL and GroES purified by affinity chromatography. a, Radioactivity profile. b, Absorption (A) at 214 nm, indicative of peptide bonds. c, Absorption at 297 nm, indicative of $8N_3$ -nucleotide. That part of the elution profile containing peptides is shown.

METHODS. GroEL (2 mg) and GroES (0.3 mg) were incubated for 10 min at 4 °C in 2 ml 20 mM MOPS/KOH, pH 7.2, 100 mM KCl, 5 mM magnesium acetate with 2 μ M [α - ^{32}P] $8N_3$ -ATP and then irradiated (Fig. 1 legend). The procedure was repeated after addition of 300 μ M unlabelled $8N_3$ -ATP. Samples were precipitated with 7% trichloroacetic acid and the pellets resuspended in 170 μ l 2 M urea, 75 mM NH_4CO_3 , pH 8.5, containing 7.5 mg trypsin (Sigma). More trypsin (7.5 mg) was added after 3 h and 6 h and incubation continued overnight at 23 °C. The reaction mixture was diluted 2-fold with 50 mM $NH_4(OOCCH_3)$, adjusted to pH 5.8 with acetic acid, and applied onto a 1.2-ml Al(III)-chelate column, prepared as described¹⁵. After successive washes with 15 ml 50 mM $NH_4(OOCCH_3)$, pH 5.8, and 10 ml 50 mM $NH_4(OOCCH_3)$, pH 5.8/0.5 M NaCl, bound peptides were eluted with 6 ml 5 mM KH_2PO_4 /0.1 M $NH_4(OOCCH_3)$, pH 8.0. Fractions containing radioactivity were pooled and 25% of the total eluate was further analysed on a reversed-phase Vydac 214TP52 C4 (2.1 $\times$ 250 mm) column using a 3.5–70% acetonitrile gradient in 0.1% trifluoroacetic acid. The HPLC system was as described²³, with a diode-array detector allowing real-time monitoring of absorbance at 214 nm and 297 nm. Radioactivity was determined by liquid scintillation counting. Fractions 38–43 (1) and 46–48 (2) were pooled as indicated and the peptides contained in these fractions were sequenced^{24,25} and analysed by matrix-assisted laser desorption mass spectrometry^{26,27}. All amino acids of the peptides shown in Fig. 3a could be identified unequivocally, except for Tyr 477 in GroEL and Tyr 71 in GroES, which were apparently modified by the nucleotide analogue.

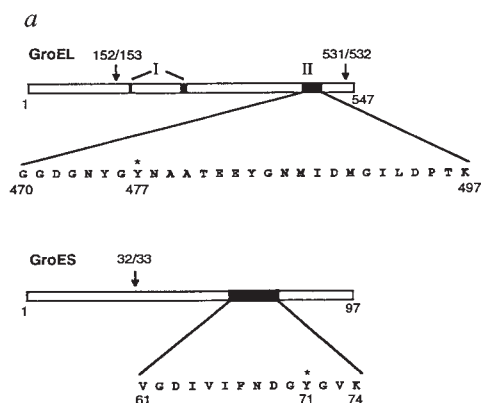
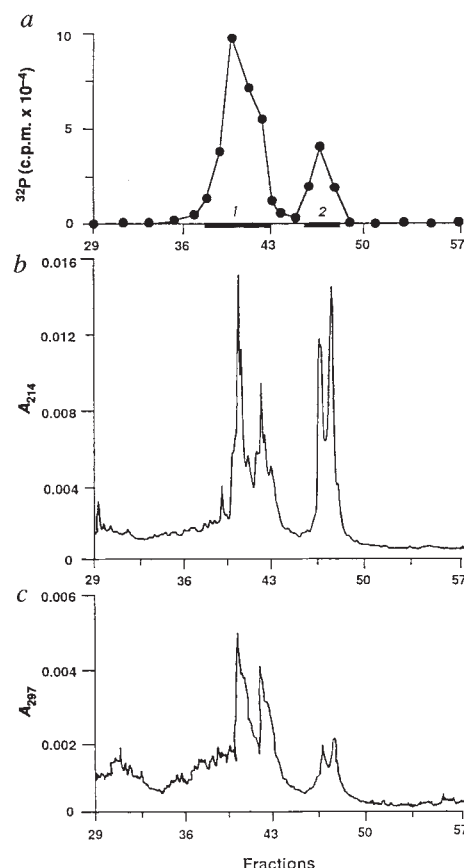


FIG. 3 a, Amino-acid sequences of $8N_3$ -nucleotide-modified peptides and their locations within the GroEL and GroES proteins. I, A putative consensus motif for binding of the triphosphate group of ATP; II, $8N_3$ -nucleotide-labelled peptide of GroEL. Asterisk denotes the modified Tyr residues in GroEL and GroES, respectively; arrows indicate the positions of the proteinase K cleavage sites in GroEL (residues 531/532) and GroES (residues 32/33) which are protected in the GroEL–GroES complex, and the cleavage site in N-ethylmaleimide(NEM)-treated GroEL (residues 152/153) (Fig. 4). b, Comparison of the sequence regions

b

GroES

GroEL

<i>E. coli</i>	68NDGYGVK..	<i>E. coli</i>	474NYGYNA..
<i>B. abortus</i>	70GK-WSGT..	<i>B. abortus</i>	474TFGYNTA..
<i>B. subtilis</i>	66SK-YAGT..	<i>B. burgdorferi</i>	473GLGFDAS..
<i>C. burnetii</i>	68GK-YAGT..	<i>B. napus1</i>	475KFGYNAA..
<i>C. perfringens</i>	67SK-YAGT..	<i>B. napus2</i>	474EIGYNAM..
<i>C. psittaci</i>	75DK-YAGQ..	<i>B. subtilis</i>	471GVGFNAA..
<i>C. trachomatis</i>	75DK-YSGQ..	<i>C. burnetii</i>	479NYGYNA..
<i>L. micdadei</i>	68GK-YSGT..	<i>C. perfringens</i>	472GVGFDA..
<i>M. tuberculosis</i>	71SK-YGGT..	<i>C. psittaci</i>	474SEGYDAL..
<i>P. aeruginosa</i>	68GP-YSGS..	<i>C. trachomatis</i>	473NEGYDAL..
<i>R. norvegicus</i>	73PE-YGGT..	<i>H. ducreyi</i>	471NFGYNAT..
<i>R. tsutsugamushi</i>	66TK-WAGT..	<i>H. sapiens</i>	473EVGYDAM..
<i>S. albus</i>	75SK-YGGT..	<i>L. micdadei</i>	471NFGFNAA..
<i>S. oleracea1</i>	75SK-YTGT..	<i>L. pneumophila</i>	472NYGFNAA..
<i>S. oleracea2</i>	174SK-YAGN..	<i>M. leprae</i>	472GHGLNAA..
<i>S. sp. 6301</i>	76SK-YAGT..	<i>M. musculus</i>	473EVGYDAM..
<i>Thermophilic P3</i>	67SK-YAGT..	<i>M. tuberculosis</i>	472GHGLNAQ..
		<i>R. communis</i>	463EIGYNAM..
		<i>R. norvegicus</i>	473EVGYDAM..
		<i>R. tsutsugamushi</i>	471NRGFDA..
		<i>S. albus</i>	476Q-GFNAA..
		<i>S. cerevisiae</i>	474DKGYDAS..
		<i>S. sp. 7942</i>	476E-GYDAS..
		<i>T. aestivum</i>	473EMGYNAM..
		<i>Thermophilic P3</i>	471GIGFNAA..

surrounding the $8N_3$ -ATP-modified Tyr residues (asterisks) in known GroEL and GroES homologues. *S. oleracea* 1 and 2 represent the first and second halves of the 'fused' chloroplast chaperonin 10. *B. napus* 1 and 2 denote the α - and β -subunits of the chloroplast chaperonin 60, respectively; for the GroEL homologues of *R. communis* and *T. aestivum*, the chloroplast α -subunits of chaperonin 60 are shown. Protein sequences and sequence alignments were obtained from the EMBL and SWISSPROT databases. Secondary structure predictions for GroEL and GroES were performed using the PHD program²⁸.

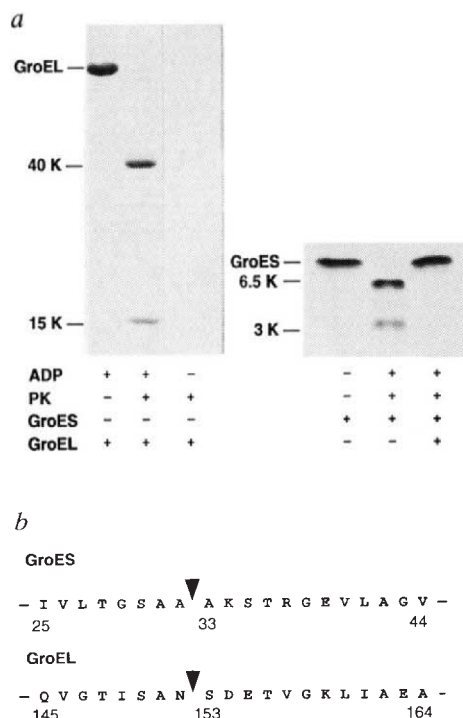


FIG. 4 *a*, Formation of GroEL and GroES domains by limited proteolysis. *b*, Amino-acid sequences around the protease cleavage sites (arrowheads).

METHODS. *a*, Left panel: GroEL (0.35 μ M) was incubated for 10 min at 23 °C in 20 mM Tris, pH 8.0, containing 5 mM NEM. The reaction then received 2 mM dithiothreitol and the pH was adjusted to 7.2 with 0.1 vol 0.6 M Tris, pH 6.8. When indicated, 5 mM magnesium acetate or 1 mM ADP were added. After 5 min, samples were cooled to 4 °C and incubated for 20 min in the absence or presence of 120 μ g ml⁻¹ proteinase K (PK) as indicated. Proteinase K action was stopped by adding 1 mM phenylmethanesulphonyl fluoride. Right panel: GroES alone (1.2 μ M) or GroES (1.2 μ M) in a complex with GroEL (1.5 μ M), formed in the presence of 1 mM Mg-ADP, was incubated for 5 min at 4 °C in buffer A (Fig. 1 legend) containing 20 μ g ml⁻¹ proteinase K where indicated. Protease treatment was stopped by addition of 1 mM phenylmethanesulphonyl fluoride. All reactions were analysed by SDS-PAGE and Coomassie blue staining. GroES fragments were separated according to ref. 29. *b*, The 3K and 6.5K fragments of GroES, and the 15K and 40K fragments of GroEL were N-terminally sequenced after electroblotting onto polyvinylidene difluoride membranes²⁶. The 3K and 15K fragments were found to start with the N termini, and the 6.5K and 40K fragments with residues 33 and 153 of GroES and GroEL, respectively.

shows that the labelled tyrosines as well as their surrounding amino-acid residues are highly conserved (Fig. 3*b*). In both proteins, secondary structure prediction locates the critical tyrosines in a non-ordered environment.

Azido-ATP labelling indicated that the carboxy-terminal third of the GroEL sequence participates in binding the purine ring of the nucleotide. This region of GroEL may also be involved in making contact with GroES, consistent with the observation that binding of GroES protects the interacting GroEL subunits against proteolytic cleavage of the 16 C-terminal residues⁴. In an attempt to assign the 8N₃-labelled tyrosine residue to a distinct nucleotide-binding domain of GroEL, we found that after treatment with *N*-ethylmaleimide and in the presence of nucleotide, GroEL could be cleaved by proteinase K into an N-terminal 15K, and a ~40K domain (Fig. 4*a*) beginning at residue 153 (Fig. 4*b*). In the absence of nucleotide, the protein was completely digested. Non-alkylated GroEL, on the other hand, is

resistant to proteinase K when nucleotide is bound, except for the truncation at the C terminus. Apparently modification with *N*-ethylmaleimide renders a flexible region near residue 153 accessible to protease. Interestingly, the 40K domain also contains a putative nucleotide-binding consensus sequence for the triphosphate group^{18,19} (residues 164–172 and 241–251; Fig. 3*a*) which, together with the region around Tyr 477, may form the nucleotide-binding pocket of GroEL.

GroES could also be fragmented by proteinase K into an *N*-terminal fragment of *M_r* ~3K and a C-terminal segment of ~6.5K (starting at residue 33) (Fig. 4), which was found to be crosslinked to 8N₃-nucleotide. Interestingly, the proteinase K cleavage site between residues 32/33 becomes inaccessible when GroES is bound to GroEL (Fig. 4). As recently detected by NMR spectroscopy, complex formation between GroES and GroEL is accompanied by the immobilization of a flexible loop in GroES between residues 16 and 31 (ref. 20). Binding of GroES also leads to protease protection of the 16-residue C-terminal segment in the subunits of the GroEL ring that binds to GroES⁹.

Considering that GroES has no detectable ATP hydrolytic activity, what could be the significance of its nucleotide-binding capability? At physiological concentrations of nucleotides, GroES will bind predominantly ATP (not shown). Formation of the GroEL–GroES complex is dependent on nucleotide^{4,7,12,20} and is stimulated when ATP binds to GroEL before ATP hydrolysis⁸. Given the relatively low affinity of GroEL for ATP, ATP binding by GroES could play a part in increasing the probability that all seven ATP sites in the interacting GroEL toroid are occupied simultaneously. This would be critical in achieving highly cooperative ATP hydrolysis to release bound substrate protein for folding. When a GroEL–GroES complex, formed in the presence of [α -³²P]8N₃-ATP, was removed from free nucleotide by gel filtration, nucleotide could be crosslinked only to GroEL⁹. In contrast, both GroES and GroEL were labelled when the free nucleotide was not removed (Fig. 1*A*). These results are consistent with the possibility that GroES donates ATP to GroEL, thus increasing the cooperativity of ATP binding. □

Received 9 September; accepted 12 October 1993.

- Ellis, R. J. & van der Vies, S. M. *A. Rev. Biochem.* **60**, 327–347 (1991).
- Gething, M. J. & Sambrook, J. *Nature* **355**, 33–45 (1992).
- Hendrick, J. P. & Hartl, F.-U. *A. Rev. Biochem.* **62**, 349–384 (1993).
- Langer, T., Pfeifer, G., Martin, J., Baumeister, W. & Hartl, F.-U. *EMBO J.* **11**, 4757–4765 (1992).
- Braig, K., Simon, M., Furuya, F., Hainfeld, J. F. & Horwich, A. L. *Proc. natn. Acad. Sci. U.S.A.* **90**, 3978–3982 (1993).
- Saibil, H. R., Dong, Z., Wood, S. & auf der Mauer, A. *Nature* **353**, 25–26 (1991).
- Saibil, H. R. et al. *Curr. Biol.* **3**, 265–273 (1993).
- Bochkareva, E. S., Lissin, N. M., Flynn, G. C., Rothman, J. E. & Girshovich, A. S. *J. biol. Chem.* **267**, 6796–6800 (1992).
- Martin, J., Mayhew, M., Langer, T. & Hartl, F.-U. *Nature* **366**, 228–233 (1993).
- Chandrasekhar, G. N., Tilly, K., Woolford, C., Hendrix, R. & Georgopoulos, C. *J. biol. Chem.* **261**, 12414–12419 (1986).
- Martin, J. et al. *Nature* **352**, 36–42 (1991).
- Viitanen, P. V. et al. *Biochemistry* **29**, 5665–5671 (1990).
- Jackson, G. S. et al. *Biochemistry* **32**, 2554–2563 (1993).
- Gray, T. & Fersht, A. R. *FEBS Lett.* **292**, 254–258 (1991).
- Shoemaker, M. T. & Haley, B. E. *Biochemistry* **32**, 1883–1890 (1993).
- Holleman, M., Runswick, M. J., Fearnley, I. M. & Walker, J. E. *J. biol. Chem.* **258**, 9307–9313 (1983).
- Knight, K. L. & McEntee, K. J. *J. biol. Chem.* **260**, 10185–10191 (1985).
- Martel, R., Cloney, L. P., Pelcher, L. E. & Hemmingsen, S. M. *Gene* **94**, 181–187 (1990).
- Fry, D. C., Kuby, S. A. & Mildvan, A. S. *Proc. natn. Acad. Sci. U.S.A.* **83**, 907–911 (1986).
- Landry, S. J., Zeilstra-Ryalls, J., Fayet, O., Georgopoulos, C. & Gierasch, L. M. *Nature* **364**, 255–258 (1993).
- Fayet, O., Louran, J. M. & Georgopoulos, C. *Molec. gen. Genet.* **202**, 435–445 (1986).
- Shlomai, J. & Kornberg, A. *J. biol. Chem.* **255**, 6789–6793 (1980).
- Tempst, P., Link, A. J., Riviere, L. R., Fleming, M. & Elicone, C. *Electrophoresis* **11**, 537–553 (1990).
- Tempst, P. & Riviere, L. R. *Analyt. Biochem.* **183**, 290–300 (1989).
- Erdjument-Bromage, H., Geromanos, S., Chodera, A. & Tempst, P. in *Techniques in Protein Chemistry IV* (ed. Angeletti, R. H.) 419–426 (Academic, San Diego, 1993).
- Beavis, R. C. & Chait, B. T. *Rapid Commun. Mass Spectrom.* **3**, 233–237 (1989).
- Geromanos, S., Casteels, P., Elicone, C., Powell, M. & Tempst, P. in *Techniques in Protein Chemistry V* (ed. Crabb, J. W.) (Academic, San Diego, in the press).
- Rost, B., Schneider, R. & Sander, C. *Trends biochem. Sci.* **18**, 120–123 (1993).
- Schägger, H. & von Jagow, G. *Analyt. Biochem.* **166**, 368–379 (1987).

ACKNOWLEDGEMENTS. We thank L. Lacomis for technical assistance. J.M. is supported by a fellowship of the Dr. Mildred Scheel-Stiftung für Krebsforschung.

Nonlinear amplification by calcium-dependent chloride channels in olfactory receptor cells

Graeme Lowe & Geoffrey H. Gold

Monell Chemical Senses Center, 3500 Market Street, Philadelphia, Pennsylvania 19104-3308, USA

THE sense of smell is highly evolved in mammals, allowing discrimination between a vast number of odorants, with detection thresholds as low as 10^{-17} M (ref. 1). Although several features of mammalian olfactory transduction have been revealed by biochemical and molecular biological studies²⁻¹¹, the odorant-induced membrane current has remained elusive. In amphibians this current is mediated by cyclic-nucleotide-gated channels¹²⁻¹⁵, which depolarize the cell by Na^+ and Ca^{2+} influx^{16,17} and consequent Cl^- efflux through Ca^{2+} -dependent Cl^- channels^{18,19}. The Cl^- current may be absent in mammals, however, because its proposed role is linked to the aquatic habitat of amphibians¹⁸. Here we show that the transduction current in rat olfactory receptor cells is initiated by cyclic-nucleotide-gated channels. The Cl^- current is also present and endows the transduction current with a steep sigmoidal dependence on cyclic AMP concentration in both rat and in an amphibian, indicating a new function for the Cl^- channel: nonlinear amplification of the transduction signal, whereby suprathreshold responses are boosted relative to basal transduction noise.

Odorant- and cAMP-induced currents were recorded in the

whole-cell mode from solitary rat olfactory receptor cells. The odorants used (Fig. 1 legend) stimulate adenylyl cyclase selectively²⁰, so the phospholipase C pathway²¹ should not contribute to our measurements. Figure 1a shows the current induced by odorant pulses at a series of membrane potentials. As in amphibians^{15,16,22}, this current is inward at negative potentials, reverses near 0 mV (+10 mV for this cell; mean $\pm$ s.d. = $+0.4 \pm 5.0$ mV for 8 cells), and shows little rectification. Odorant responses from rat also resembled those from amphibians in their waveforms and amplitudes (up to 950 pA at -70 mV), although their duration was usually about one half as long.

Pulses of a cyclic nucleotide phosphodiesterase inhibitor, IBMX, were applied to increase internal cAMP concentration as a result of basal adenylyl cyclase activity. The IBMX-induced current (Fig. 1b) exhibited the same voltage-dependence as the transduction current (Fig. 1a), suggesting that the odorant responses were mediated by a cAMP-activated conductance. The shorter latency of the response to IBMX (Fig. 1b, inset) suggests a direct action on phosphodiesterase, rather than activation of odorant receptor proteins. The initial slopes of the IBMX-induced currents were comparable with those of odorant responses, indicating that basal adenylyl cyclase activity can be as large as odorant-stimulated cyclase activity. This is consistent with the high basal cyclase activity observed in isolated rat olfactory cilia²³, but contrasts with the very low basal activity of a cloned rat olfactory adenylyl cyclase expressed in a mammalian cell line⁵.

Figure 2a shows the current induced by flash photolysis of caged cAMP²³ and that induced by odorant stimulation in the same cell. The time course of the photolysis response (trace 1: latency, 12 ms; duration at half peak amplitude, 160 ms) was much more rapid than that of the odorant response (trace 2:

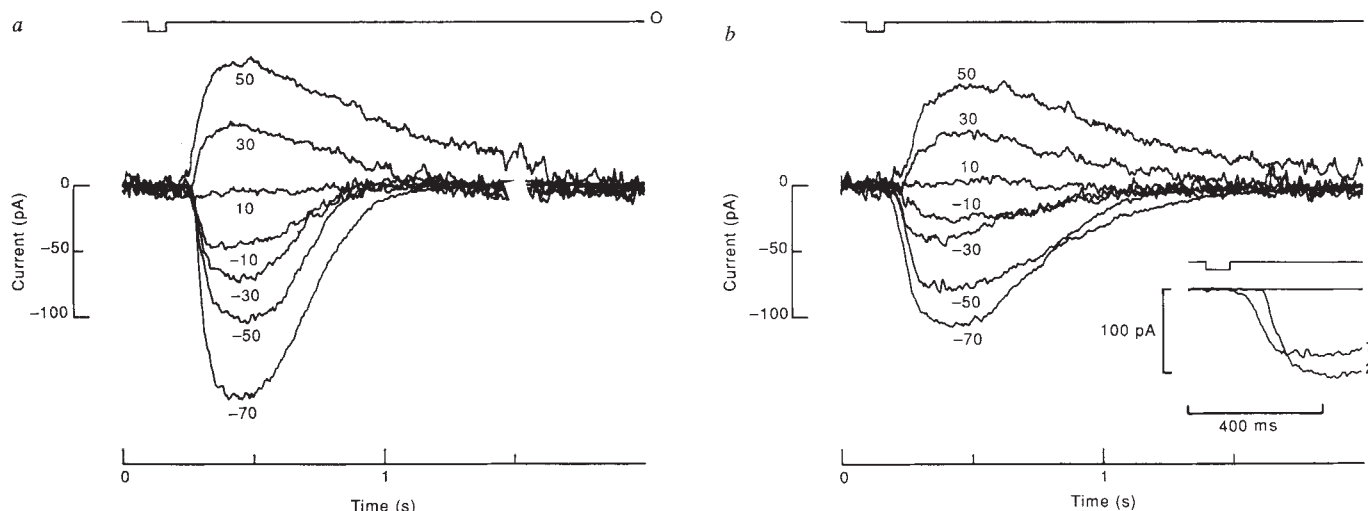


FIG. 1 Whole-cell currents from a rat olfactory receptor cell induced by brief application of a, odorants, and b, 2.5 mM IBMX (3-isobutyl-1-methyl xanthine) at a series of holding potentials (values in mV are indicated next to each trace). The stimulus parameters (ejection pressure and pulse duration) were held constant as the holding potential was changed. Stimulus trace 0, odorant; stimulus trace 1, IBMX. Inset: an IBMX response (trace 1) and an odorant response (trace 2) at -50 mV, plotted on an expanded timescale to reveal the difference in latency. The fraction of cells responding to odorants was 14% (22/161) and 40% (38/96) to IBMX.

METHODS. Solitary rat olfactory receptor cells were obtained by dissociating the olfactory epithelia of adult male Sprague-Dawley rats in low calcium/trypsin as described³⁰. Olfactory receptor cells were identified by the presence of cilia. Currents were recorded in the whole-cell configuration (holding potential -50 mV unless specified otherwise). The odorant solution was a 100-fold dilution in rat saline of saturated solu-

tions of each of the following: β -ionone, (-)-menthone, amyl salicylate, isoamyl acetate and 2-hexylpyridine. The solution containing caged cAMP (100 μ M 4,5-dimethoxy-2-nitrobenzyl adenosine-3',5'-cyclic monophosphate (Molecular Probes, Oregon)) was prepared and used as described¹⁵. For solutions that were applied by pressure ejection from a micropipette (100 kPa unless indicated otherwise), the concentration inside the pipette is given, although the concentration that enveloped the cell must have been considerably lower²². The extracellular solution contained (mM): NaCl 135, NaHCO_3 1, NaH_2PO_4 1.2, KCl 4.4, MgSO_4 1.2, CaCl_2 2.6, D-glucose 16, Na-HEPES 5, and 0.001% phenol red (sodium salt), pH 7.2. Whole-cell pipette solutions contained (mM): KCl 90, potassium aspartate 50, MgCl_2 5, NaHCO_3 1, MgSO_4 1, K_2EGTA 0.1, HEPES (potassium salt) 5, Na_2ATP 5, Na_3GTP 1, and 0.001% phenol red (potassium salt), pH 7.2. All experiments were done at 23 °C.

latency, 195 ms; duration, 945 ms), indicating that cAMP acted directly on the ciliary cyclic-nucleotide-gated conductance²⁴, and that channel gating kinetics do not determine the time course of the transduction current, contrary to a recent proposal²⁵.

Further evidence that transduction in rat is mediated by cyclic-nucleotide-gated channels was provided by observing the effect of an odorant-induced current on a simultaneous cAMP-induced current generated by photolysis of caged cAMP. In salamander, photolysis responses are enhanced during small odorant-induced currents¹⁵, which is evidence that the odorant response is mediated by cAMP. Such nonlinear summation was also observed in rat (Fig. 2*b*), with potent enhancement of photolysis responses by a weak odorant stimulus: the amplitude of the smallest photolysis response (10 pA, trace 1) was increased 25-fold by an odorant-induced current of only 20 pA (trace 2).

The nonlinear relation between membrane current and cAMP

concentration was measured directly by varying the flash intensity (Fig. 2*c*): the amplitudes of the resulting family of photolysis responses were fitted by a Hill equation with cooperativity 4.7 (Fig. 2*d*); an even higher value (6.9) was obtained for another cell. Such high cooperativities were also seen in salamander (data not shown) and in the odorant concentration dependence of the transduction current in salamander²⁶, indicating that the nonlinearity of transduction resides primarily in current generation, not in cAMP metabolism^{6,15}. The large difference between the cooperativities of the cAMP-induced current in intact cells, and the cyclic-nucleotide-gated conductance in excised patches ($n \sim 1.5$; ref. 24), is due to the Cl^- component of the whole-cell current^{18,19}, as will be shown below.

We demonstrated a Cl^- contribution to the transduction current in rat by measuring the voltage dependence of the current in a low Cl^- solution. As shown in amphibians¹⁸, this condition

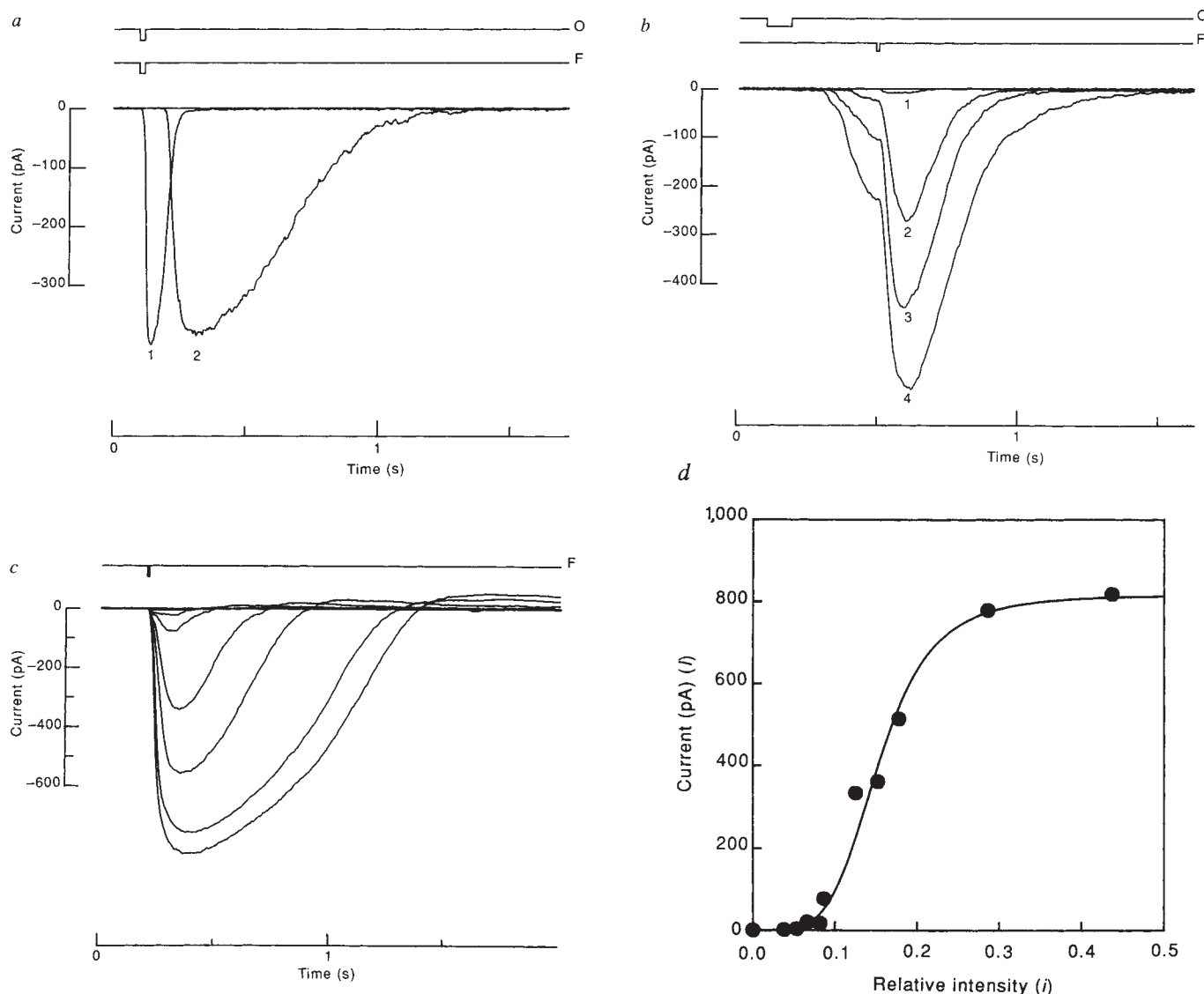


FIG. 2 Whole-cell currents induced by odorants and by flash photolysis of caged cAMP. *a*, Comparison between the time course of a photolysis response (trace 1) and an odorant response (trace 2). The responses were recorded at different times from the same cell, but were plotted to superimpose the two stimuli on the time axis (flash duration, 40 ms (stimulus trace F); odorant pulse duration, 40 ms (stimulus trace O)). The flash intensity and odorant pulse pressure were adjusted to produce responses of similar peak amplitude. *b*, Nonlinear summation of a photolysis response and odorant responses of different amplitudes. Identical flashes (duration 10 ms, intensity attenuated by 0.132) were timed to occur at the peaks of responses induced by 90-ms odorant pulses

of increasing pressure (traces 2–4, 35–140 kPa; trace 1, flash only). *c*, Responses to flashes of various intensities (duration 10 ms; light attenuations: 0.053, 0.066, 0.086, 0.125, 0.178, 0.217, 0.286). A linear relation between flash intensity and cAMP production was confirmed by observing that a 10-ms flash and a 20-ms flash at half the intensity produced an identical photolysis response. *d*, Plot of the peak amplitudes of the photolysis-induced currents from *c*, measured 100 ms after the flash, as a function of relative flash intensity. Data were fitted by the Hill equation, $I = I_m i^n / (i^n + K^n)$, with parameters $I_m = 817$ pA, $K = 0.15$, $n = 4.7$.

shifts the reversal potential of the Cl^- current positive with respect to that of the cyclic-nucleotide-gated cationic current, resulting in a biphasic current at +10 mV (Fig. 3a). The fraction of the total transduction current that was carried by Cl^- in physiological saline was estimated from the attenuation of the current by the Cl^- channel blocker niflumic acid¹⁹. Application of niflumic acid during an odorant response caused an immediate reduction in membrane current, followed by a slower increase

in current (Fig. 3b). Another Cl^- channel blocker, SITS, had a similar effect. We attribute the initial reduction to block of the Cl^- current, because niflumic acid and SITS are both specific blockers of the Cl^- channels in amphibian receptor cells^{18,19}. The secondary current may be due to an increase in cAMP concentration because its voltage dependence was similar to that of the transduction current (data not shown). To determine the amplitude and time course of the Cl^- component free of the

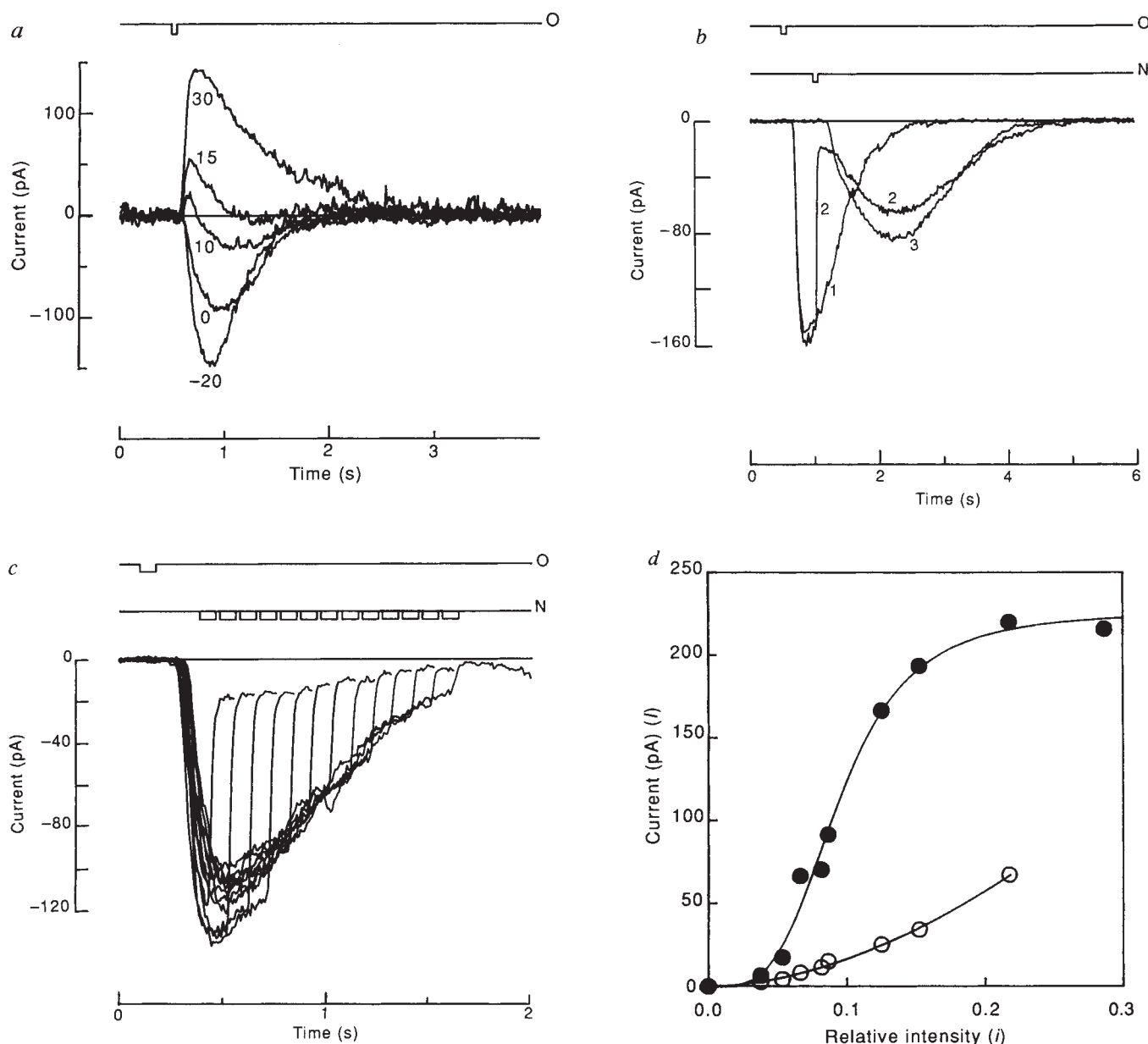


FIG. 3 Chloride component of the transduction current. *a*, Whole-cell currents induced by identical odorant pulses at a series of holding potentials in a low Cl^- bathing solution (steady-state current subtracted). Odorant pulse duration was 50 ms; holding potentials are indicated in mV next to each trace; the external solution contained 9.6 mM chloride (135 mM glutamate⁻ was substituted for 135 mM Cl^- in the extracellular solution). *b*, Effect of brief application of niflumic acid (500 μM) on the odorant-induced current: trace 1, odorant response; trace 2, odorant response with a niflumic acid pulse applied at the peak; trace 3, response to niflumic acid. *c*, A family of traces in which niflumic acid pulses (stimulus trace N) occurred at various times during responses to identical odorant pulses (stimulus trace O). Each trace was truncated ~ 100 ms after channel block to prevent the secondary

rise from obscuring other traces. Odorant and niflumic acid pulses lasted 80 ms. *d*, Plot of the photolysis-induced current, I , as a function of flash intensity, i , in the presence (circles) or absence (filled circles) of 5 mM SITS (4-acetamido-4'-isothiocyanatostilbene-2,2'-disulphonic acid). Flash duration, 20 ms; intensity is relative to that of the unattenuated light source¹⁵. Data were obtained from a newt (*Cynops pyrrhogaster*) olfactory receptor cell as described in ref. 30. To avoid prolonged exposure to the chloride channel-blocker, SITS was applied by pressure ejection from a micropipette for 3 s, starting 1 s before each flash. Current amplitudes were measured 100 ms after the flash. The following functions were used to fit the data: circles: $I = ki^n$, $k = 1014$ pA, $n = 1.78$; filled circles: $I = I_m i^n / (i^n + K^n)$, $I_m = 226$ pA, $K = 0.095$, $n = 3.7$.

secondary effect, pulses of niflumic acid were applied at different times during an odorant response. The family of traces in Fig. 3c reveals that, as in amphibians¹⁸, the time course of the Cl⁻ current (the rapid decrement in current caused by niflumic acid) is similar to that of the cationic current (that remaining after the decrement). In this experiment, however, the Cl⁻ current constituted a larger fraction of the transduction current (~85%) than has been reported in amphibians (~40%)¹⁸.

Additional information about the relative amplitudes of the cationic and Cl⁻ currents was obtained by observing the effect of a Cl⁻ channel blocker on the intensity dependence of the photolysis response. This experiment was performed using newt receptor cells in the continuous presence of SITS, because the secondary effect of SITS was absent in amphibians. Our results (Fig. 3d) show that the total current (filled circles) increased more steeply with cAMP concentration than did the cationic current alone (circles), and that, as in rat (Fig. 3b, c), the Cl⁻ component (filled circles minus circles) contributed a large fraction of the total current (up to 85% between 20 and 170 pA). Thus, the Cl⁻ current is a nonlinear function of cAMP concentration, and increases the apparent cooperativity of the cAMP-induced current from ~1.5 in excised patches²⁴ to ~3.5 in an intact cell (Fig. 3d). The same conclusion should apply to rat cells because of the parallel difference between cAMP cooperativity in excised patches ($n=1.5$; ref. 27) and in the intact cell ($n=4.7$; Fig. 2d). The nonlinear activation of the Cl⁻ current may be due to both the nonlinear Ca²⁺ dependence of the Cl⁻ conductance²⁸ and Ca²⁺ buffering, which would introduce a threshold into activation of the conductance.

The function of the Cl⁻ current may be related to the strong nonlinearity that it confers on the cAMP-induced current. This would result in amplification of suprathreshold responses relative to basal transduction noise. Minimizing basal transduction

noise may be important in olfactory receptor cells because, at threshold, currents of only a few pA are sufficient to generate action potentials²⁹. Such nonlinear amplification may be important in other cells whose output is encoded by action potentials. □

Received 11 June; accepted 14 October 1993.

1. Passe, D. H. & Walker, J. C. *Neurosci. Biobehav. Rev.* **9**, 431–467 (1985).
2. Shirley, S. G., Robinson, J., Dickinson, K., Aujla, R. & Dodd, G. H. *Biochem. J.* **240**, 605–607 (1986).
3. Sklar, P. B., Anholt, R. R. H. & Snyder, S. H. *J. biol. Chem.* **261**, 15538–15543 (1986).
4. Jones, D. T. & Reed, R. R. *Science* **244**, 790–795 (1989).
5. Bakalyar, H. A. & Reed, R. R. *Science* **250**, 1403–1406 (1990).
6. Boekhoff, I., Tareilus, E., Strotmann, J. & Breer, H. *EMBO J.* **9**, 2453–2458 (1990).
7. Buck, L. & Axel, R. *Cell* **65**, 175–187 (1991).
8. Ronnett, G. V., Parfitt, D. J., Hester, L. D. & Snyder, S. H. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2366–2369 (1991).
9. Restrepo, D. et al. *Am J. Physiol.* **263**, C667–673 (1992).
10. Ben-Arie, N., Khen, M. & Lancet, D. *Biochem. J.* **292**, 379–384 (1993).
11. Raming, K. et al. *Nature* **361**, 353–356 (1993).
12. Kurahashi, T. *J. Physiol.* **430**, 355–371 (1990).
13. Firestein, S., Zufall, F. & Shepherd, G. M. J. *Neurosci.* **11**, 3565–3572 (1991).
14. Frings, S. & Lindemann, B. *J. gen. Physiol.* **97**, 1–16 (1991).
15. Lowe, G. & Gold, G. H. *J. Physiol.* **462**, 175–196 (1993).
16. Kurahashi, T. *J. Physiol.* **419**, 177–192 (1989).
17. Kurahashi, T. & Shibuya, T. *Brain Res.* **515**, 261–268 (1990).
18. Kurahashi, T. & Yau, K.-W. *Nature* **363**, 71–74 (1993).
19. Kleene, S. J. *Neuron* **11**, 123–132 (1993).
20. Breer, H. & Boekhoff, I. *Chem. Senses* **16**, 19–29 (1991).
21. Hugue, T. & Bruch, R. C. *Biochem. biophys. Res. commun.* **137**, 36–42 (1986).
22. Firestein, S. & Werblin, F. *Science* **244**, 79–82 (1989).
23. Nerbonne, J. M., Richard, S., Nargeot, J. & Lester, H. A. *Nature* **310**, 74–76 (1984).
24. Nakamura, T. & Gold, G. H. *Nature* **325**, 442–444 (1987).
25. Zufall, F., Hatt, H. & Firestein, S. *Proc. natn. Acad. Sci. U.S.A.* **90**, 935–939 (1993).
26. Firestein, S., Picco, C. & Menini, A. *J. Physiol.* **468**, 1–10 (1993).
27. Frings, S., Lynch, J. W. & Lindemann, B. *J. gen. Physiol.* **100**, 45–67 (1992).
28. Kleene, S. J. & Gesteland, R. C. *J. Neurosci.* **11**, 3624–3629 (1991).
29. Lynch, J. W. & Barry, P. H. *Biophys. J.* **55**, 755–768 (1989).
30. Lowe, G. & Gold, G. H. *J. Physiol.* **442**, 147–168 (1991).

ACKNOWLEDGEMENTS. We thank S. J. Kleene for suggesting the use of niflumic acid, and A. J. Hudspeth, T. Kurahashi and L. M. Masukawa for their comments on the manuscript. This work was supported by a grant from the USNIH.

GUIDE TO AUTHORS

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without display items, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. High priority cannot be given to pre-submission enquiries; in urgent cases they can be made in the form of a one-page fax. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are processed from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose.

Once a manuscript is accepted for publication, contributors will receive proofs in about 4 weeks. *Nature's* staff will edit manuscripts with a view to brevity and clarity, so contributors should check proofs carefully. Manuscripts are generally published 2–3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review Articles survey recent developments in a field. Most are commissioned but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Progress articles review particularly topical developments for a nonspecialist readership. They do not exceed 4 pages in length. Suggestions may be made to the Reviews Coordinator in the form of a brief synopsis.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50–80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few short subheadings.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research.

News and Views articles inform nonspecialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the News and Views Editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and four copies are required, each accompanied by artwork. If photographs are included, five sets of originals are required; for line drawings, one set of originals and four good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Three copies of relevant manuscripts in press or submitted for publication elsewhere should be included with submitted manuscripts, clearly marked as such. Five copies of revised and resubmitted manuscripts, labelled with their manuscript numbers, are required, together with five copies of a letter detailing the changes made.

Titles are brief and simple. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs.

Colour artwork. A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963–65). The first and last page numbers are included; reference to books includes publisher, place and date.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used in the text.

Acknowledgements are brief and appear after the reference list; grant and contribution numbers are not allowed.

Supplementary information is material relevant to Articles or Letters which cannot, for lack of space, be published in full, but which is available from *Nature* on request.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1234 National Press Building, Washington, DC 20045, USA. A telephone and fax number should be included. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax.